APPLICATION-BASED PACKET ROUTING IN VEHICULAR NETWORKS

by

Kaveh Shafiee

B.Sc., University of Tehran, 2005
M.Sc., Sharif University of Technology, 2007

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES
(Electrical and Computer Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

July 2012

© Kaveh Shafiee, 2012
Abstract

A wide range of vehicular applications require packet routing mechanisms and protocols for efficient, reliable and robust delivery of data packets over vehicles, from a single source or from multiple sources to either a specific destination or multiple destinations in a specific region. All vehicular applications, e.g., non-safety applications and safety applications, have their own specific challenges, considerations and Quality of Service (QoS) requirements which are different from those of other vehicular applications. Hence, we believe that the design of packet routing mechanisms and protocols for each vehicular application should be application-specific.

In non-safety applications, to access the backhaul network through the infrastructure on the roadside, so-called Road-side Units (RSUs), since many parts of the network may not be directly covered by RSUs, appropriate routing protocols need to be designed and employed. In this thesis, we developed a single-technology routing protocol for Vehicular Ad hoc Networks (VANETs), Connectivity-aware Minimum-delay Geographic Routing (CMGR), which adapts well to continuously changing network status in such networks. As the next step we studied how packet routing mechanisms and protocols should be adapted to heterogeneous environments. In this regard, we developed optimal Vertical Hand-Off (VHO) strategies for vehicular heterogeneous networks when RSUs directly cover all parts of the vehicular network under study. Next, we turned our attention to the case where some parts of the network are not directly covered by any RSU and proposed a Hybrid Multi-Technology Routing (HMTR) protocol to consider different combinations of wireless technologies in intermediate hops when establishing routes from vehicular end-users to RSUs.
In safety applications the notification of hazardous situations needs to be sent from the hazard-detecting vehicle to every other vehicle in the neighborhood, so-called data dissemination. Fully ad hoc data dissemination mechanisms have gained more acceptance due to their robustness and avoiding the excessive costs of infrastructure deployment and maintenance. In this regard, one of the main challenges is to overcome the packet delivery failures at intersections in the ad hoc manner. In this regard, we developed a fully ad hoc data dissemination mechanism, Enhanced Intersection-mode Data Dissemination (EIDD), which provides reliable packet delivery at intersections.
Preface

The following publications describe the work completed in this thesis. In some cases, the conference papers contain materials overlapping with the journal papers. All the chapters are based on these papers co-authored with Dr. Victor C.M. Leung, who also supervised the research. One of the papers related to Chapter 3 was also co-authored with Dr. Alireza Attar. I am the principal author of all the publications.

Journal Papers Accepted/Published:

Conference Papers Published:


Table of Contents

Abstract............................................................................................................................................... ii
Preface.................................................................................................................................................. iv
Table of Contents .................................................................................................................................. vi
List of Tables ......................................................................................................................................... x
List of Figures ....................................................................................................................................... xi
List of Abbreviations .......................................................................................................................... xv
Acknowledgements ............................................................................................................................. xviii
Dedication ............................................................................................................................................. xix

Chapter 1: Introduction ........................................................................................................................... 1
  1.1 Overview of Vehicular Applications (Trends, Challenges and Needs) ........................................ 3
    1.1.1 Packet routing for non-safety applications ................................................................. 3
    1.1.2 Packet routing for safety applications ........................................................................ 7
  1.2 Summary of Main Contributions ............................................................................................... 8
  1.3 Thesis Organization ..................................................................................................................... 13

Chapter 2: Connectivity-aware Minimum-delay Geographic Routing with Vehicle Tracking in VANETs .................................................................................................................. 14
  2.1 Introduction ................................................................................................................................. 14
  2.2 Connectivity-aware Minimum-delay Geographic Routing ................................................... 17
    2.2.1 Assumptions and system model .................................................................................... 17
    2.2.2 CMGR protocol operations ......................................................................................... 18
    2.2.3 Route selection logic ................................................................................................. 19
### Chapter 2: Performance Limitation Scenarios and Solutions

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3.1 Vehicle tracking mechanism for moving destinations</td>
<td>24</td>
</tr>
<tr>
<td>2.3.2 Packet forwarding decision scheme at sparse junctions</td>
<td>28</td>
</tr>
</tbody>
</table>

### Chapter 2: Performance Evaluations

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4.1 Simulation settings</td>
<td>32</td>
</tr>
<tr>
<td>2.4.2 Simulation results</td>
<td>35</td>
</tr>
</tbody>
</table>

### Chapter 2: Summary

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5 Summary</td>
<td>40</td>
</tr>
</tbody>
</table>

### Chapter 3: Optimal Distributed Vertical Handoff Strategies in Vehicular

#### Heterogeneous Networks

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Introduction</td>
<td>43</td>
</tr>
<tr>
<td>3.2 Problem Formulation</td>
<td>48</td>
</tr>
<tr>
<td>3.3 VHO Decision-making Algorithm with Fixed AP Inter-distances</td>
<td>50</td>
</tr>
<tr>
<td>3.3.1 Cost-minimization approach</td>
<td>51</td>
</tr>
<tr>
<td>3.3.2 Transmission-time minimization approach</td>
<td>53</td>
</tr>
<tr>
<td>3.4 VHO Decision-making Algorithm with Statistical AP Inter-distances</td>
<td>54</td>
</tr>
<tr>
<td>3.4.1 Cost-minimization approach</td>
<td>58</td>
</tr>
<tr>
<td>3.4.2 Transmission-time minimization approach</td>
<td>58</td>
</tr>
<tr>
<td>3.5 VHO Decision-making with Enabled V2V Mode</td>
<td>58</td>
</tr>
<tr>
<td>3.5.1 Ad hoc networking delay analysis</td>
<td>59</td>
</tr>
<tr>
<td>3.5.2 Calculation of cost and transmission time</td>
<td>62</td>
</tr>
<tr>
<td>3.5.3 WLAN plus cellular plus ad hoc</td>
<td>63</td>
</tr>
<tr>
<td>3.6 Performance Evaluation</td>
<td>64</td>
</tr>
<tr>
<td>3.7 Summary</td>
<td>72</td>
</tr>
</tbody>
</table>
Chapter 4: Hybrid Multi-Technology Routing in Heterogeneous Vehicular Networks

4.1 Introduction ......................................................................................................................... 76

4.2 Network Topology .............................................................................................................. 81

4.2.1 Assumptions ..................................................................................................................... 81

4.2.2 The topology .................................................................................................................... 81

4.3 Hybrid Multi-Technology Routing ...................................................................................... 85

4.3.1 Dissemination of a request packet ................................................................................... 85

4.3.2 Route selection .................................................................................................................. 86

4.3.3 Returning a reply packet ................................................................................................. 87

4.4 Hybrid Multi-Technology Routing ...................................................................................... 89

4.4.1 Route selection logic ....................................................................................................... 89

4.4.1.1 QoS and (or) budget filtering ....................................................................................... 90

4.4.1.2 Candidate route selection ......................................................................................... 92

4.4.2 Link stability logic .......................................................................................................... 95

4.5 Route Connectivity ............................................................................................................ 97

4.6 Performance Evaluation ..................................................................................................... 103

4.6.1 Preliminaries ..................................................................................................................... 103

4.6.2 Simulation settings ......................................................................................................... 104

4.6.3 Simulation results .......................................................................................................... 107

4.7 Summary ............................................................................................................................ 114

Chapter 5: A Reliable Robust Fully Ad Hoc Data Dissemination Mechanism for Vehicular Networks .................................................................................................................. 116

5.1 Introduction ......................................................................................................................... 116
5.2 Directional mode data dissemination mechanisms ................................................. 120
5.3 Intersection mode data dissemination mechanisms ............................................... 125
5.4 Proposed enhanced intersection mode data dissemination mechanism .................... 128
5.5 Summary ............................................................................................................. 133

Chapter 6: Concluding Remarks and Possible Future Directions ................................. 135
6.1 Summary of accomplished work ......................................................................... 135
6.2 Future directions ................................................................................................. 138

Bibliography ............................................................................................................ 141
List of Tables

Table 2.1  Mobility-related and wireless communications-related parameters .................. 34
Table 2.2  Challenges and features of CMGR ................................................................. 42
Table 3.1  Variables and their definitions ........................................................................ 50
Table 3.2  Mobility-related and wireless communication-related parameters used in the simulation .................................................................................................................... 65
Table 3.3  Challenges and features in the proposed VHO strategies .............................. 74
Table 4.1  Mobility-related parameters .............................................................................. 106
Table 4.2  WLAN-related parameters ................................................................................. 106
Table 4.3  WiMAX-related parameters ............................................................................... 106
Table 4.4  Transmission costs for 1 attachment point ....................................................... 110
Table 4.5  Transmission costs for 2 attachment points ..................................................... 112
Table 5.1  Mobility-related and wireless communication-related parameters used in the simulation .................................................................................................................... 122
List of Figures

Fig. 2.1  An example illustrating CMGR and the vehicle tracking mechanism .......................... 27
Fig. 2.2  A typical scenario using packet forwarding decision scheme ................................. 31
Fig. 2.3  Pseudo code of CMGR .................................................................................................. 31
Fig. 2.4  The simulation street layout .......................................................................................... 34
Fig. 2.5 (a) Packet delivery ratio for maximum delay of 1 min. and 1 GW, and (b) Packet delivery delay for maximum delay of 1 min. and 1 GW ......................................................... 37
Fig. 2.6 (a) Packet delivery ratio for maximum delay of 10 sec. and 1 GW, and (b) Packet delivery delay ratio for maximum delay of 10 sec. and 1 GW .......................................................... 37
Fig. 2.7  Ratio of dropped data packets for maximum delay of 1 min. and 1 GW ..................... 39
Fig. 2.8  Ratio of dropped data packets for maximum delay of 10 sec. and 1 GW ................. 39
Fig. 2.9  Packet delivery delay for maximum delay of 1 min. and different number of GWs ............................................................................................................................................. 40
Fig. 2.10 (a) Packet delivery ratio for maximum delay of 10 sec. and different number of GWs, and (b) Packet delivery delay for maximum delay of 10 sec. and different number of GWs ............................................................................................................................................. 40
Fig. 3.1  The reference model ...................................................................................................... 44
Fig. 3.2  A vehicular network with fixed inter-distance APs .................................................... 52
Fig. 3.3  Random inter-distance APs in a vehicular scenario .................................................... 56
Fig. 3.4  Two consecutive APs with distance $d=A+W$ ............................................................ 61
Fig. 3.5  Comparison of transmission time performance in the proposed algorithms, when $b_t=1$ Gb .......................................................... 66

Fig. 3.6  The cost of communication using different VHO policies for the scenario similar to Fig. 3.5 .................................................................................................................. 67

Fig. 3.7  The required number of VHO for the proposed vehicular communication strategies .................................................................................................................. 67

Fig. 3.8  Upper and lower delay bounds for $d_{hop}$ [66] .......................................................... 68

Fig. 3.9  Ad hoc delay as a function of inter-distance of APs ................................................. 70

Fig. 3.10 Comparison of transmission time performance in the proposed algorithms .... 72

Fig. 3.11 The cost of communication using different VHO policies for the scenario similar to Fig. 3.10 .................................................................................................................. 72

Fig. 4.1 Network topology ........................................................................................................ 83

Fig. 4.2 A typical description of a route ...................................................................................... 89

Fig. 4.3 Leaving borders and other parameters for the vehicles in subsection 4.4.2 .......... 97

Fig. 4.4 Different flows of vehicle arriving in the range of Vehicle $i$ ................................. 99

Fig. 4.5 Distances $R$ ahead and $R$ behind vehicle $i$ modeled as 2 $M/D/\infty$ queues ........... 101

Fig. 4.6 (a) Packet Delivery Ratio for maximum delay of 1 min. and 1 attachment point, and (b) Packet Delivery Delay for maximum delay of 1 min. and 1 attachment point .... 109

Fig. 4.7 (a) Packet Delivery Ratio for maximum delay of 10 sec. and 1 attachment point, and (b) Packet Delivery Delay for maximum delay of 10 sec. and 1 attachment point .... 109

Fig. 4.8 (a) Packet Delivery Ratio for maximum delay of 1 min. and 2 attachment points, and (b) Packet Delivery Delay for maximum delay of 1 min. and 2 attachment points.... 112
Fig. 4.9  (a) Packet Delivery Ratio for maximum delay of 10 sec. and 2 attachment points, and (b) Packet Delivery Delay for maximum delay of 10 sec. and 2 attachment points...... 112

Fig. 4.10  (a) Packet Delivery Ratio for maximum delay of 1 min. and different percentage of WiMAX-enabled vehicles, and (b) Packet Delivery Delay for maximum delay of 1 min. and different percentage of WiMAX-enabled vehicles ........................................... 113

Fig. 4.11  (a) Packet Delivery Ratio for maximum delay of 10 sec. and different percentage of WiMAX-enabled vehicles, and (b) Packet Delivery Delay for maximum delay of 10 sec. and different percentage of WiMAX-enabled vehicles ........................................... 113

Fig. 5.1  A chain accident example (a) no wireless data dissemination, and (b) wireless enabled data dissemination ................................................................. 118

Fig. 5.2  A typical failure scenario............................................................................. 120

Fig. 5.3  The comparison between UMB and DP-IB when employed on a straight highway (for 20 vehs/km per lane and 2 beacons per second in DP-IB) ....................................... 124

Fig. 5.4  The latency decreases when the beacon interval in DP-IB increases, but still remains greater than UMB (for avg. speed of 35 m/sec) ............................................. 124

Fig. 5.5  (a) A forwards the message to B which is out of the intersection, (b) A forwards the message to B which is inside the intersection, but leaves the intersection before C arrives at its transmission range, and (c) A had better forward the packet to entering node B rather than leaving node C ......................................................................................... 128

Fig. 5.6  How EIDD overcomes the failure scenario in Fig. 5.2 .............................. 130

Fig. 5.7  The primary road intersected by three secondary roads .......................... 131

Fig. 5.8  The delivery ratio of EIDD is considerably higher than AMB and approximately 100% ............................................................................................................. 132
Fig. 5.9  EIDD and AMB have approximately similar delays........................................ 132
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>Authentication, Authorization and Accounting</td>
</tr>
<tr>
<td>ABS</td>
<td>Antilock Brake System</td>
</tr>
<tr>
<td>ACK</td>
<td>Acknowledgment</td>
</tr>
<tr>
<td>AMB</td>
<td>Ad hoc Multi-hop Broadcast</td>
</tr>
<tr>
<td>AMC</td>
<td>Adaptive Modulation and Coding</td>
</tr>
<tr>
<td>AP</td>
<td>Access Point</td>
</tr>
<tr>
<td>A-STAR</td>
<td>Anchor-based Street and Traffic Aware Routing</td>
</tr>
<tr>
<td>BE</td>
<td>Best Effort</td>
</tr>
<tr>
<td>BS</td>
<td>Base Station</td>
</tr>
<tr>
<td>CBR</td>
<td>Constant Bit Rate</td>
</tr>
<tr>
<td>CMGR</td>
<td>Connectivity-aware Minimum-delay Geographic Routing</td>
</tr>
<tr>
<td>CTB</td>
<td>Clear-To-Broadcast</td>
</tr>
<tr>
<td>CTS</td>
<td>Clear-To-Send</td>
</tr>
<tr>
<td>DP-IB</td>
<td>Data Pouring with Intersection Buffering</td>
</tr>
<tr>
<td>DSRC</td>
<td>Dedicated Short Range Communications</td>
</tr>
<tr>
<td>EEBL</td>
<td>Electronic Emergency Break Light</td>
</tr>
<tr>
<td>EIDDD</td>
<td>Enhanced Intersection-mode Data Dissemination</td>
</tr>
<tr>
<td>FCC</td>
<td>Federal Communications Commission</td>
</tr>
<tr>
<td>FCW</td>
<td>Forward Collision Warning</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
</tr>
<tr>
<td>GW</td>
<td>Gateway</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>HMTR</td>
<td>Hybrid Multi-Technology Routing</td>
</tr>
<tr>
<td>IP</td>
<td>Internet protocol</td>
</tr>
<tr>
<td>ITS</td>
<td>Intelligent Transport System</td>
</tr>
<tr>
<td>JS</td>
<td>Junction Sequence</td>
</tr>
<tr>
<td>LLT</td>
<td>Link Lifetime</td>
</tr>
<tr>
<td>LOS</td>
<td>Line-Of-Sight</td>
</tr>
<tr>
<td>LSS</td>
<td>Location Service Server</td>
</tr>
<tr>
<td>MAC</td>
<td>Medium Access Control</td>
</tr>
<tr>
<td>MADM</td>
<td>Multi-Attribute Decision-Making</td>
</tr>
<tr>
<td>MOPR</td>
<td>Movement Prediction-based Routing</td>
</tr>
<tr>
<td>MORA</td>
<td>Movement-based Routing Algorithm</td>
</tr>
<tr>
<td>MOVE</td>
<td>Mobility model generator for Vehicular networks</td>
</tr>
<tr>
<td>MURU</td>
<td>Multi-hop Routing protocol for Urban vehicular ad hoc networks</td>
</tr>
<tr>
<td>NLOS</td>
<td>Non-Line-Of-Sight</td>
</tr>
<tr>
<td>NS-2</td>
<td>Network Simulator 2</td>
</tr>
<tr>
<td>OFDMA</td>
<td>Orthogonal Frequency-Division Multiple Access</td>
</tr>
<tr>
<td>PBR</td>
<td>Prediction-Based Routing</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>RD</td>
<td>Route Discovery</td>
</tr>
<tr>
<td>RR</td>
<td>Route Reply</td>
</tr>
<tr>
<td>RSS</td>
<td>Received Signal Strength</td>
</tr>
<tr>
<td>RSU</td>
<td>Roadside Unit</td>
</tr>
<tr>
<td>RTB</td>
<td>Request-To-Broadcast</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------------</td>
</tr>
<tr>
<td>RTS</td>
<td>Request-To-Send</td>
</tr>
<tr>
<td>SAR</td>
<td>Spatially-Aware Routing</td>
</tr>
<tr>
<td>SUMO</td>
<td>Simulation of Urban Mobility</td>
</tr>
<tr>
<td>TT</td>
<td>Trip Time</td>
</tr>
<tr>
<td>UMB</td>
<td>Urban Multi-hop Broadcast</td>
</tr>
<tr>
<td>V2R</td>
<td>Vehicle-to-Roadside</td>
</tr>
<tr>
<td>V2V</td>
<td>Vehicle-to-Vehicle</td>
</tr>
<tr>
<td>VADD</td>
<td>Vehicle-Assisted Data Delivery</td>
</tr>
<tr>
<td>VANET</td>
<td>Vehicular Ad hoc Network</td>
</tr>
<tr>
<td>VHN</td>
<td>Vehicular Heterogeneous Network</td>
</tr>
<tr>
<td>VHO</td>
<td>Vertical Hand-Off</td>
</tr>
<tr>
<td>WiMAX</td>
<td>Wireless interoperability for Microwave Access</td>
</tr>
<tr>
<td>WLAN</td>
<td>Wireless Local Area Network</td>
</tr>
</tbody>
</table>
Acknowledgements

I would like to express my deep gratitude to my supervisor Dr. Victor C.M. Leung for his insightful guidance, invaluable support and constant encouragement from the early stages of my research until the very end. Without his wisdom and constructive suggestions, this work would not be possible. I would like to thank my supervisory committee members Dr. Garland Chow and Dr. Dave Michelson, and departmental examination committee member Dr. Vincent Wong for their invaluable comments and for their support and help during my PhD program.

I am also thankful to my fellow colleagues and friends at the Department of Electrical and Computer Engineering at the University of British Columbia who inspired me along the way with their constructive comments and continuous support, especially Mr. Pooyan Abouzar, Mr. Amr Alasaad, Mr. Pedram Ataee, Dr. Alireza Attar, Mr. SeyedAli Hosseininezhad, and Mr. Peyman TalebiFard.

Lastly, I would like to thank my parents and my sister Kiana for their support during these years.

This study was supported in part by the Canadian Network of Centers of Excellence Program through AUTO21 NCE, the Canadian Natural Sciences and Engineering Research Council (NSERC) and industrial and government partners through NSERC-DIVA Strategic Research Network, the Nokia University Relations Program, the BC-China ICSD Program, and the University of British Columbia International Partial Tuition Scholarship.
Dedication

With love, to my parents.
Chapter 1: Introduction

As automotive technology progresses into the 21st century, the need for wireless communications between vehicles becomes apparent. In this regard, the Federal Communications Commission (FCC) of the United States allocated 75 MHz spectrum to Dedicated Short Range Communications (DSRC) [1] at 5.9 GHz frequency band to Vehicle-to-Vehicle (V2V) and Vehicle-to-Roadside (V2R) communications. DSRC was standardized in IEEE 802.11p standard [2], a variant of IEEE 802.11a standard adjusted for low overhead operations, which supports data rates up to 6 Mbps. On the other hand, since a regulation requiring DSRC in every car for the enhancement of public safety appears possible, car manufacturers as one of the key players in the value chain are interested in equipping their products with access network interfaces, so-called wireless radios, complying with DSRC 802.11p standard. This is expected to provide vehicles with the opportunity to wirelessly communicate with each other and with the roadside infrastructure. Since vehicular nodes can use multi-hop packet forwarding to communicate with each other, vehicular networks are commonly referred to as Vehicular Ad hoc Networks (VANETs).

A wide range of vehicular applications require some sort of mechanisms and protocols for efficient, reliable and robust delivery of data packets over vehicles, so-called packet routing, from a source node or multiple source nodes to either a specific destination node or multiple destinations in a specific region or in larger surroundings. What makes the packet routing in vehicular networks more challenging than other readily existing networks are the intrinsic characteristics of such networks, i.e., highly variable network topologies [3], frequent network fragmentation at lower densities [4] and the fact that movements of vehicles are constrained to pre-defined roads with specific speed limits [5]. The general trend
of research on vehicular networking has been based on the modification and extension of readily available non-vehicular protocols and mechanisms to vehicular scenarios to adopt the particular characteristics of vehicular networking environments. Examples of such a trend will be introduced in the future chapters. Since vehicular networking is expected to enable a wide range of new applications and services that did not exist before, e.g., non-safety applications and safety applications, some of the requirements of these new applications have not been considered in the design of older mechanisms and protocols, thereby making the extension of non-vehicular protocols to vehicular scenarios extremely inefficient and difficult. That is why many researchers have recently come to the conclusion that instead of extending older mechanisms and protocols, new packet routing mechanisms and protocols should be designed from scratch exclusive to every new application with respect to the specific challenges, considerations and Quality of Service (QoS) requirements of the application of concern, i.e., application-specific. We take the same approach in our thesis.

In other words, in the present thesis, we particularly develop packet routing mechanisms and protocols from the viewpoint of application needs, i.e., we propose application-specific packet routing mechanisms and protocols for various vehicular applications that carefully take the challenges, considerations and specific requirements of each vehicular application into account. As we will observe, one of the main advantages of our proposed routing solutions, which stems from the application-specific design approach, is that they can adapt to and show acceptable performance in a wide spectrum of situations and scenarios, ranging from dense and sparse networks in rural and urban areas to slow and fast moving vehicles with highly varying topologies and complex road structures. In the evaluation of every
proposed packet routing protocol or mechanism, the tradeoffs between performance metrics, connectivity measures and budget constraints are taken into consideration.

The rest of this chapter is organized as follows. In Section 1.1 different types of vehicular applications are described and the specific packet routing mechanisms and protocols required for each of the applications as well as their design challenges are explained. Then we identify the specific problems that are yet to be addressed for any given application and explain how solving these problems contribute to the overall goal of our thesis. Section 1.2 provides a summary of the solutions that we provide in our thesis and discusses the significance and novelty of the proposed mechanisms and protocols. Finally, the organization of the thesis is described in Section 1.3.

1.1 Overview of Vehicular Applications (Trends, Challenges and Needs)

1.1.1 Packet routing for non-safety applications

An important type of vehicular applications are non-safety applications, also referred to as comfort or entertainment applications, ranging from interactive games, interactive communications such as video-conferencing, roadside service applications such as weather information and location and price lists of restaurants and gas-stations to Internet Protocol (IP)-based applications such as email, web browsing, content delivery, file download, gaming services, IP telephony and multimedia streaming [6]. In non-safety applications vehicular end-users need to access the backhaul network through some infrastructure that provides the interfaces to the backhaul network. The infrastructure mostly consists of Roadside Units (RSUs) or Gateways (GWs) or attachment points, e.g., Base Stations (BSs) in the case of Wireless interoperability for Microwave Access (WiMAX) or cellular networks, or Access Points (APs) in the case of Wireless Local Area Networks (WLANs)
such as IEEE 802.11 a/b/g/p WLANs. Since many parts of the network may not be directly covered by RSUs, appropriate multi-hop routing protocols need to be designed and employed to establish efficient routes from vehicular end-users to RSUs, i.e., from one single source to one single destination, possibly through multiple vehicles.

Several challenges should be considered in the design of routing protocols for VANETs. The main design challenges are robustness to fast topology changes, awareness of the street layout and considering real-time connectivity information in the network. If the design of the routing protocol is not adaptable to the dynamic nature of vehicular environments, the established route may be unstable and it is likely that the vehicular nodes that are forming the route leave each others’ transmission ranges and the route fails even before any data is sent. On the other hand, if the routing protocol does not consider the real-time connectivity information in the network, the packets could be forwarded towards disconnected streets where no next-hop vehicle can be found and as a result the packets are dropped and the route fails. In this respect, in Chapter 2 we develop a routing protocol for VANETs which is adaptable to highly variable network topologies and considers the real-time connectivity information of the network.

Most of the existing routing protocols for VANETs including the protocol that we propose in Chapter 2 are based on a single wireless access technology, i.e., all the vehicular nodes involved in the packet routing use the same type of wireless radio and wireless standard. We call this type of routing protocols single-technology routing protocols. In recent years, the extensive development of wireless communication systems has resulted in the availability of several wireless access networks, each employing a different wireless access technology, at any geographic area, e.g., cellular networks, WiMAX networks and
WLANs. On the other hand, nowadays vehicular users are being equipped with multiple wireless radios. Vehicular networks will particularly benefit from such a rich set of connectivity alternatives, so-called heterogeneous network, as the heterogeneous network is expected to meet higher levels of QoS demands of vehicular users compared to a single wireless access network. On the other hand, heterogeneous networks can simultaneously address multiple requests each from a different vehicular end-user regarding a different vehicular application, whereas using a single access network can hardly, if ever, address the needs of all different vehicular applications at the same time.

In Vehicular Heterogeneous Networks (VHNs), when RSUs, each belonging to a different access network, directly cover all parts of the network under study, the main challenge is to make appropriate decisions on which access network to use for receiving service among different possible connectivity alternatives at any point in time. The decisions on which access network among available alternatives to establish a connection or switch to, commonly referred to as Vertical Hand-Off (VHO), can depend on both user and network preferences and should be made based on performance, availability or economical reasons. Even though many VHO algorithms have been proposed for this purpose [7-12], when it comes to vehicular environments, most of the readily available VHO algorithms turn out to be inefficient and ineffective. This inefficiency stems from the fact that in the design of these VHO algorithms the mobility models of users including their movement trajectories and their velocities are often neglected. More specifically, since vehicles travel at high speeds, it is very likely that handing off from a wide-coverage network to a newly emerged local-coverage network may be followed by another VHO back to the original network immediately afterwards resulting in too many VHOs that imposes VHO processing loads and
signalling overhead to the network. In Chapter 3 we develop optimal mobility-aware VHO strategies for vehicular heterogeneous networks to make appropriate decisions on which access network to use for receiving service at any point in time.

On the other hand, when some parts of the vehicular heterogeneous network under study are not directly covered by any RSU, as mentioned earlier, routing protocols are needed to establish efficient routes to the most appropriate RSU in a larger neighbourhood. However, due to the heterogeneous nature of the network under study, the routing protocols to be used in this environment should consider different combinations of wireless technologies in intermediate hops and find the most appropriate set of intermediate hops. We call this type of routing protocols multi-technology routing protocols. Very few papers have investigated the problem of multi-technology routing in heterogeneous networks [13-15]. The number of papers on multi-technology routing for vehicular networks is even less. The common assumption in all of these studies is that there exists an access network with global coverage, such as the cellular or WiMAX network, to which end-users can connect as a back-up alternative at any time at any location, particularly when the access networks with smaller coverage areas such as WLANs are unavailable. Since the size of vehicular networks may become extremely large in practice, having a global back-up network may not be realistic and all of the access networks in the scenario under study should be considered as independent connectivity alternatives as opposed to global or local alternatives.

Furthermore, in highly dynamic vehicular environments with heterogeneous connectivity alternatives, the wireless technologies employed over links should preferably be adapted to the stability level of the links. In other words, when long-range wireless technologies such as WiMAX networks are used for packet forwarding, vehicular nodes stay in each others’
ranges for a longer time yielding more stable links compared to short-range wireless technologies such as WLAN. None of the previous studies considers the stability of links in the process of selecting the appropriate wireless technology over a given link. In Chapter 4 we propose a hybrid multi-technology routing protocol that considers different available access networks as independent connectivity options and characterizes intermediate links as stable and variable based on their expected lifetimes and uses position-based and topology-based routing over highly variable links and more stable links, respectively. In position-based routing, also called geographic routing, protocols packet-forwarding decisions are made on-the-fly based on the positions and movement attributes of current immediate neighbours of the packet forwarding vehicle [16, 17]. So, in position-based routing the positions of the forwarding nodes form the route as opposed to the nodes themselves, which is the case in topology-based routing. Hence, position-based routing is a better choice for highly varying topologies, such as vehicular networks. On the other hand, in topology-based routing the route is composed of a number of intermediate nodes, and the route is established before sending the data packets.

1.1.2 Packet routing for safety applications

Another important type of vehicular applications is safety applications that are aimed to improve the safety of drivers and passengers of vehicles and pedestrians on roads. In safety applications, to avoid impending collisions, the notification of hazardous situations, e.g., when airbag deploys in case of an accident or when Antilock Brake System (ABS) is activated in case of slippery road conditions [18, 19], needs to be sent from the hazard-detecting vehicle to every other vehicle present in the neighbourhood, i.e., from a single source to multiple destinations. This type of packet routing is called data dissemination.
Clearly, one of the critical performance metrics in the design of data dissemination mechanisms is packet delivery latency that should be kept low enough to make sure drivers have enough time to react to the safety notifications that they receive. Another critical performance design metric is reliability which means the delivery of safety data, i.e., hazard notifications, to all the vehicles in the area of interest should be guaranteed. Many data dissemination mechanisms have been previously proposed by the research community [20-29]. Fully ad hoc data dissemination mechanisms have gained more acceptance due to their robustness, i.e., not relying on infrastructure, and avoiding the excessive costs of infrastructure deployment and maintenance. In this regard, one of the main challenges is to overcome the packet delivery failures at intersections. This could happen when the hazard notification leaves the intersection region without making sure that it has been forwarded towards all the out-going roads. In Chapter 5 we propose a fully ad hoc data dissemination mechanism that provides reliable packet delivery at intersections.

1.2 Summary of Main Contributions

In the present thesis, for various vehicular applications we propose application-specific packet routing mechanisms and protocols that carefully take the challenges, considerations and specific requirements of each vehicular application into account in their designs. The proposed packet routing mechanisms and protocols are presented in four chapters as follows:

- As mentioned in Subsection 1.1.1, two main challenges should be dealt with in the design of any plausible single-technology routing protocol for vehicular networks: adaptability to fast topology changes and avoiding disconnected routes. Even though many routing protocols have been proposed to deal with the dynamic nature of vehicular environments by selecting more stable routes or more stable next hops [16,
30-36], the connectivity-awareness has only been studied in very few papers [37, 38]. However, these papers do not use accurate enough metrics to assess the level of connectivity in the intermediate streets. The inaccuracy stems from the fact that the metrics employed are based on stochastic models and average values stored in offline databases, and consequently do not reflect the latest connectivity levels of the streets on the roadmap. On the other hand, an effective routing protocol for vehicular networks should have good performance regardless of the status of the network. In other words, even though in sparse situations the main challenge is to take the connectivity of intermediate streets into account to maximize the chance of packet reception, when the network is dense and connected in most parts, the main design challenge is to minimize the packet delivery delay by selecting non-congested routes that have a sufficient level of connectivity. With these two objectives in mind, in Chapter 2 we develop a Connectivity-aware Minimum-delay Geographic Routing (CMGR) protocol for VANETs which is adaptable to the continuously changing status of the network and considers real-time connectivity information of the network. The proposed protocol is compared with two plausible geographic connectivity-aware routing protocols for VANETs and the obtained results show that under the specific conditions considered in the simulations, the packet delivery ratio of CMGR is approximately 25% better than its peers for high vehicle densities and goes up to 900% better for low vehicle densities. Parts of this chapter have been included in a published journal article [90], and a published conference paper [91].

• Even though many VHO decision-making algorithms have been proposed for heterogeneous networks in the literature [7-12, 39-43], they might be inefficient and
ineffective in VHNs with highly dynamic topologies and highly variable environment conditions. This inefficiency stems from the fact that in the design of these algorithms the mobility models of vehicular users, particularly the fact that they travel through several access networks in a short span of time due to their high speeds, are often neglected. In such algorithms, it is very likely that handing off from a wide-coverage network to a newly emerged local-coverage network may be followed by another VHO back to the original network immediately afterwards resulting in too many VHOs. Therefore, the network is overloaded with heavy VHO signalling traffic that causes additional costs and longer transfer times. Very few studies have exploited the knowledge of mobility patterns of vehicular users in a VHO decision-making algorithm [43]. However, all of these studies are based on a centralized location server. The server is aware of the coverage areas of all the access points in the network and also receives periodic location reports from every vehicle in the network. Based on this information, it makes VHO decisions at the beginning of discrete time intervals by predicting the mobility of vehicles. Clearly, in addition to the high costs of infrastructure deployment and maintenance, such a centralized scheme is also not scalable in VHNs and incurs heavy processing loads in the core network and heavy packet traffic caused by vehicle location reporting. In Chapter 3 we propose an optimal VHO decision-making algorithm, which is fully distributed in the sense that vehicular users make the VHO decisions for themselves rather than location servers in the core network. Also, the proposed algorithm is event-activated, and thus continuous-time in the sense that vehicles continuously select optimal access networks based on their mobility patterns at any point in time as opposed to relying on discrete time intervals. The presented results
provide insightful guidelines for optimal VHO decision-making based on the characteristics of the network as well as the user mobility profile. Parts of this chapter have been included in a published journal article [92].

- Although some papers have investigated the issue of multi-technology routing in heterogeneous environments, they make the general assumption that the access network with larger coverage area provides blanket coverage over the investigated locale, thereby making it possible for the end-users to directly connect to this anywhere-available access network at any time. As the size of vehicular networks may become extremely large in practice, considering such back-up network may not be realistic. On the other hand, topology-based routing protocols are fairly unstable in vehicular environments and consequently not scalable beyond a few hops even when the stability of routes is incorporated in the process of route establishment [32-34]. Also, position-based routing protocols are resilient to highly varying topologies, but the forwarding decisions are local and without considering real-time network conditions in terms of connectivity and congestion in other parts of the network. Even when the connectivity information is considered in packet-forwarding decisions [37, 38], the connectivity information used is offline and not reflecting the up-to-date conditions in the parts of the network that are going to be visited. In Chapter 4 we propose a Hybrid Multi-Technology Routing (HMTR) protocol for VHNs that integrates the advantages of topology-based and position-based routing into a unified scheme. In the heterogeneous network under study, all access networks regardless of the size of their coverage areas are used as independent connectivity alternatives. HMTR adapts the wireless technologies employed over links to the stability level of the links by characterizing
intermediate links as stable and variable based on their expected lifetimes and uses position-based and topology-based routing over highly variable links and more stable links, respectively. Parts of this chapter have been included in an accepted journal article [93], and a published conference paper [94].

• Many solutions for disseminating data from a source vehicle to a large number of vehicles in the neighbourhood of interest have been previously proposed by the research community. In order to deploy a data dissemination mechanism for real-world provisioning of safety applications, it requires to be 100% reliable in different vehicle densities in the network for various road structures including both straight roads and intersections. Note that a reliable data dissemination mechanism should guarantee the data delivery to every single vehicle in the area of interest. Despite demonstrating reliable performance when disseminating data along straight roads, most of the previously proposed studies encounter difficulties when disseminating data around intersections. The data delivery failure at an intersection occurs when the data is being forwarded across the intersection and it leaves the intersection before it is forwarded towards all the possible out-going roads in the intersection. In Chapter 5 we propose an Enhanced Intersection-mode Data Dissemination mechanism (EIDD) to overcome delivery failure around intersections. EIDD introduces a scheme to guarantee that upon data disseminating at an intersection the data has been forwarded towards all the outgoing streets at the intersection before it leaves the intersection. The proposed protocol is compared with another plausible fully ad hoc data dissemination mechanism with intersection awareness. The obtained results show that EIDD outperforms its peer in terms of delivery ratio in different vehicle densities. Under the specific conditions
considered in the simulations, the delivery ratio of EIDD is 100% for approximately all the vehicle densities that shows its scalability and robustness to dense situations where the network becomes congested and packet collisions are most likely. Parts of this chapter have been included in a published journal article [95], and a published conference paper [96].

1.3 Thesis Organization

The remainder of the thesis is organized as follows. Our proposed single-technology connectivity-aware position-based routing protocol for VANETs, CMGR, and the mechanisms required to address the performance limitations of CMGR in special vehicular networking situations are presented in Chapter 2. Our optimal distributed event-activated VHO decision-making algorithm for vehicular heterogeneous environments in a comprehensive set of system models, pertinent to V2R communications as well as in scenarios with both V2R and V2V communications, is described in Chapter 3. We explain our proposed hybrid multi-technology routing protocol for vehicular heterogeneous environments, HMTR, and the mechanisms and logic designed for HMTR in Chapter 4. Our novel united fully ad hoc data dissemination mechanism for traffic safety applications, EIDD, which is designed to provide reliable data dissemination both along straight roads and at intersections is discussed in Chapter 5. Finally, Chapter 6 summarizes the main results, conclusions and possible directions for future research. Each of the main chapters in the thesis is self-contained and included in separate journal articles and conference papers. For any of the research problems studied in each chapter, a review of the corresponding related work for the particular problem under study is given in the chapter accordingly. The notations have been consistent throughout the thesis.
Chapter 2: Connectivity-aware Minimum-delay Geographic Routing with Vehicle Tracking in VANETs

2.1 Introduction

Many applications in VANETs, particularly non-safety applications, require the deployment of a routing protocol that is efficient, reliable and robust to frequent route disruptions. Some typical examples of these applications are Internet access, content delivery and file downloads. In these applications the vehicle in need of accessing the Internet should employ a routing protocol to establish a multi-hop connection with the most appropriate GW in its neighbouring area. GWs are the points of attachment to the backhaul network and discovering them in the neighbouring area is usually carried out in an anycasting manner. What make the design of effective routing protocols in VANETs challenging are the inherent characteristics of vehicular communications, including highly dynamic network topologies and highly variable vehicle densities.

One possibility for routing protocols in VANETs are topology-based or on-demand routing protocols in which an end-to-end route between the vehicle and the GW needs to be established before any data packet is sent. One of these protocols is Multi-hop Routing protocol for Urban vehicular ad hoc networks (MURU) [30], in which each intermediate vehicle estimates the quality of the wireless link between itself and its downlink vehicle and updates the value of a metric called the expected disconnection degree accordingly. Finally, the route with the lowest breakage probability is selected. Another topology-based routing protocol is Movement Prediction-based Routing (MOPR) [31, 32], which determines the most stable route in terms of lifetimes of the links in the route by taking the movements of vehicles into account. In most topology-based protocols new routes are only discovered when...
existing routes fail, which incurs large delays due to the reconstruction of new routes. In contrast, Prediction-Based Routing (PBR) [33] predicts the lifetimes of routes based on the mobility patterns of vehicles and discovers new routes before the existing ones fail. Another protocol that predicts route failures before they occur based on vehicular mobility patterns is proposed in [34]. In this work, to guarantee the stability of routes, vehicles with similar velocity vectors are given higher priorities in establishing routes.

Since the topology of a vehicular network can change rapidly, the amount of signalling traffic that is generated in the process of discovering an end-to-end route may be unreasonably large compared to the data traffic that actually uses the route before it fails. Furthermore, when the network is sparse, end-to-end routes may not even exist. Therefore, topology-based protocols have not gained as much popularity as position-based or geographic protocols where the next forwarding vehicles are not initially determined and packet forwarding decisions are made based on current neighbours of the packet-forwarding vehicle. One of these protocols is Movement-based Routing Algorithm (MORA) [35] in which a metric is defined on the basis of the moving directions of vehicles and their distances to the line between the source and destination nodes, and vehicles forward their data to the neighbours with the best values of the metric. In [16] the authors use the idea of MOPR to develop a geographic routing protocol. In this work, every forwarding vehicle obtains a list of its neighbours that are predicted to stay in its transmission range for at least 1 second based on their velocity vectors, and among them selects the one closest to the destination as the next hop. In these protocols there is no mechanism for avoiding topology holes caused by spatial constraints on roadmaps in vehicular environments. To consider the information on roadmaps, Spatially-Aware Routing (SAR) is proposed in [36], in which the streets and
junctions on a roadmap are mapped onto the sides and vertices of a graph and Dijkstra’s algorithm is used to find the shortest path to the destination.

All the aforementioned geographic routing protocols fail when the packets are forwarded toward a disconnected street, i.e., a street with no vehicle. The situation in which the packet-forwarding vehicle cannot find any next-hop vehicle along the route to forward the packet is called a local maximum. To overcome this issue, connectivity-aware geographic routing protocols were proposed. One of these protocols is Anchor-based Street and Traffic Aware Routing (A-STAR) [37], in which the streets with larger numbers of bus lines running along them are considered as streets with higher connectivity. Hence, based on the number of bus lines, different static weights are assigned to different streets and the Dijkstra’s algorithm is applied to compute the route with the minimum sum of weights. Another connectivity-aware protocol is Vehicle-Assisted Data Delivery (VADD) [38]. In VADD the average delays of packets on each street are computed by taking the average vehicle density and average vehicular velocity on the street into account. Then, a stochastic model is used to obtain the packet forwarding priorities for out-going roads at junctions aiming to minimize end-to-end delivery delays.

An effective routing protocol for vehicular networks should have good performance regardless of the status of the network. When the network is sparse the main challenge is to maximize the chance of reception before packets expire, by taking the connectivity of streets into account. On the other hand, when the network is dense and consequently connected in most parts, the main challenge in the design of the routing protocol is to minimize the delay by selecting non-congested routes that have a sufficient level of connectivity over time. With these two objectives in mind, we have developed our proposed routing protocol, (CMGR), as
presented in Section 2.2. The performance limitations that the proposed protocol experiences in real-world vehicular scenarios will be investigated and addressed in Section 2.3. Among the existing routing protocols only A-STAR and VADD consider the connectivity of streets when the network is sparse in their routing algorithms. Therefore, we only compare our proposal with these protocols in the performance evaluation presented in Section 2.4. Section 2.5 summarizes the chapter.

2.2 Connectivity-aware Minimum-delay Geographic Routing

2.2.1 Assumptions and system model

As in most other geographic protocols for VANETs [16, 35-38, 44], we assume that vehicles are equipped with Global Positioning System (GPS) receivers and they periodically send beacon messages reporting their positions to their neighbours. So, every vehicle can calculate the vehicle density in its immediate area. To make the beaconing more efficient, we may adopt the idea of adapting beaconing period to the vehicle densities or velocities in the neighbouring area or abrupt changes in network topology [44, 45]. However, for the sake of simplicity we assume a fixed beaconing frequency for now. Also we assume that vehicles are equipped with digital maps with detailed locations of streets and junctions. Such digital maps have already been commercialized [46].

In our system model for vehicular networks we assume that GWs, which are attachment points to the backhaul network, are arbitrarily distributed along the roadsides. Therefore, some areas in the network may not be covered by any GW. To access any point on the backhaul network, a vehicle needs to establish a connection to the backhaul network via one of the GWs in its neighbouring area. That is when the routing protocol comes into play to set up a multi-hop route between the vehicle and the GW. Since any one of the available GWs in
the neighbourhood of the vehicle can be used to attach the vehicle to the backhaul network, we may use the idea of anycasting in the system model. Due to the implementation of anycasting as part of IPv6 as in *Host Anycasting Service* [47], in this chapter we consider IPv6 as our network layer protocol.

### 2.2.2 CMGR protocol operations

Any vehicle that wants to set up a route to any GW, also called *requesting vehicle*, generates a Route Discovery (RD) message including its *identity* (ID), *location*, *velocity vector*, and the *generation time* of the message, and broadcasts it in the network. Any intermediate vehicle that receives the RD attaches its *location* to the RD before rebroadcasting it. The intended recipient of the RD is any one of the GWs in the network. Forwarding the RD message to any one of the GWs requires the implementation of anycasting in the network. Anycast addresses are the same as unicast addresses; however, in the anycast addressing scheme, the same address is assigned to more than one node. Accordingly, in our case all the GWs in the proximity of the vehicle, which are the intended recipients of the RD, share the same anycast address.

The issue of address assignment to different nodes in the network requires further studies and is outside the scope of our study. Further, to constrain the dissemination of RDs in the network, we define a *message lifetime* field in the RD, which value depends on the application that is requesting the route. When any intermediate vehicle receives the RD, it subtracts the generation time of the RD from the current time and drops the packet if the result exceeds the message lifetime.

Among all the RDs that are received at a GW for the same request but coming from different routes the GW selects the most appropriate one according to the *route selection*
logic which will be described in Subsection 2.2.3. Then, based on the locations of the intermediate vehicles included in the selected RD, the GW determines all the junctions along the route that the RD has come from as the Junction Sequence (JS). Then, the GW generates a Route Reply (RR) message comprising the JS, and the ID, location and velocity vector of the requesting vehicle already provided in the RD. The RR is sent back to the requesting vehicle along the JS by using a greedy geographic forwarding mechanism. In this mechanism any forwarding vehicle forwards the packet to the neighbour geographically closest to the next intended junction in the JS or starts carrying the packet if a local maximum occurs. After receiving the RR, the requesting vehicle will use the JS included in the RR for sending and receiving data packets to and from the GW.

2.2.3 Route selection logic

In vehicular networks, movements of vehicles are confined to streets and the widths of streets are usually much smaller than the radio transmission range of a vehicle. Therefore streets may be abstracted as line segments and the spatial movements of vehicles can be considered as one-dimensional movements along these line segments. Consequently, in this study vehicle density is measured in one dimension with the unit veh/km, rather than in two dimensions with the unit veh/km².

To take the connectivity of routes into consideration in selecting the most appropriate route, a naive approach is to select the route with the maximum value of the minimum vehicle density along the route. For this purpose, the RD rebroadcast by an intermediate vehicle includes the vehicle density in its neighbouring area, denoted by \( \rho_i \) for vehicle \( i \). Based on the periodic beacon messages that a vehicle receives, it knows the number of vehicles in its transmission range. So, by dividing the number of vehicles by the length of its
coverage area (since we are only interested in the coverage area along the line segment representing the street on which the vehicle moves) the vehicle density in its immediate neighbourhood is obtained. The route with the maximum value of the minimum vehicle density along the route is the most connected route at any point in time. However, vehicle densities are highly variable and this approach does not take their changes over time into account. In other words, at the time of decision-making the density information based on which the route is selected may be obsolete and consequently not valid. To deal with this issue, we propose the following mechanism.

Each vehicle calculates the *average expected value of the vehicle density changing rate* in its neighbouring areas, denoted by $r_i$ in the neighbouring area of vehicle $i$, and attaches it to the RD that it rebroadcasts. In order to obtain $r_i$, vehicle $i$ first computes the expected value of the density changing rate in its vicinity over a number of beaconing periods based on the beacon messages that it receives from its neighbours. This helps to reduce the effects of short-term variations of vehicle density incurred by immediate arrivals or departures of vehicles in the transmission range. By having vehicles include these expected values in their beacons, vehicle $i$ can then take the average of all the expected values of its neighbours to obtain the average expected value of the density changing rate, i.e., $r_i$. On the other hand, for any of the received RDs the GW calculates the *trip time* ($TT_j$) which is the duration of time between the generation time of the RD and the reception time of the RD at the GW, i.e., $TT_j$ along route $j$. Considering that trip times include both wireless multi-hop transmission delays and packet carrying delays in some parts caused by sparse network situations and also knowing that the GWs are connected to other GWs and servers via a high-speed backhaul network, the communication delays between the GWs are negligible compared to trip times.
Therefore, the vehicle density in the neighbouring area of vehicle $i$ when the RR gets back along route $j$, denoted by $\rho_{ija}$, can be approximated as,

$$\rho_{ija} \approx \rho_i + r_i(2T_T^j).$$

(2.1)

Note that in the above equation $\rho_i$ is the most recently calculated vehicle density in the neighbourhood of vehicle $i$ when it rebroadcasts the RD. However, $\rho_{ija}$ is the estimated vehicle density the GW predicts the neighbourhood of vehicle $i$ will have at the time the RR is sent back to its neighbourhood. The reason we consider a fixed value for $r_i$ over $2T_T^j$ is that in urban areas there is a high correlation between the current value of vehicle density changing rate and its value after maximum allowable message lifetime [48]. By changing the number of beaconing periods over which the expected values of vehicle density rates are calculated we can adjust the accuracy of the estimate in (2.1). We define connectivity along route $j$, denoted by $C_j$, as follows:

$$C_j = \min_{\text{vehicle}\in\text{route } j} (\rho_i + r_i(2T_T^j)).$$

(2.2)

Note that since the time it takes the message to get back to any intermediate vehicle is smaller than $2T_T^j$, when $r_i$ has a negative value and consequently the vehicle density is decreasing, $\rho_i + r_i(2T_T^j)$ will be a lower bound for $\rho_{ija}$. Since the minimum value is taken into account in the definition of connectivity in (2.2), having a lower bound will in turn keep the protocol on the safe side. We define $U = \{\text{route 1, route 2, ..., route } n\}$ as the set of all candidate routes. The GW selects the route, namely route $k$, with the maximum connectivity in $U$ as the most appropriate route; i.e.,

$$\text{route } k = \arg \max_{\text{route } j \in U} (C_j).$$

(2.3)

The issue with the aforementioned logic is that in dense situations it is very probable that the route with the maximum connectivity will be the route with the maximum level of
congestions, which incurs large delays and low available bandwidths. To avoid congestion, we should improve the route selection logic in (2.3) as follows

\[
route_k = \begin{cases} 
\arg \min_{route \in V} (TT_j) & \exists route \in V: 1/C_j < R \\
\arg \max_{route \in U} (C_j) & \text{otherwise}
\end{cases}, \quad (2.4)
\]

where \( R \) is the transmission range and \( V = \{route \in U, 1/C_j < R\} \). Since vehicle density has the unit of veh/km in this work, the reciprocal of the minimum average vehicle density along a route represents the maximum average distance between the vehicles on that route. The condition in (2.4) differentiates the situations where vehicles are sufficiently dense from situations where vehicles are sparse. In the dense enough situations, at least one route with relative distances of vehicles smaller than the transmission range exists, whereas in the sparse situations no such route can be found and therefore packets need to be partly carried by vehicles before they are forwarded.

In (2.4), even though \( 1/C_j \) is the maximum of average distances between vehicles on route \( j \), the fact that \( 1/C_j < R \) does not guarantee connectedness along route \( j \), particularly when the difference between \( 1/C_j \) and \( R \) is small, which means that the route is on the verge of becoming disconnected. However, in this partly connected situation the minimum delay route is most likely the most connected one, thereby maintaining the accuracy of the route selection logic for a wide range of densities. To investigate the accuracy of the route selection logic, we used global knowledge to recognize the few cases where \( 1/C_j < R \) but route \( j \) is disconnected. We observed that in all such cases the most connected route is in fact the same as the minimum delay route, which ensures the accuracy of the route selection logic. On the other hand, by employing this route selection logic we make sure that in dense
situations the route with minimum delay, which is dense enough but not congested is selected.

If the RD is received by several GWs, each of them selects the most appropriate route by using the route selection logic and sends back an RR. Upon the reception of these RRs at the requesting vehicle, it computes the connectivity ($C_j$) and trip-time ($TT_j$) of each route and selects the most appropriate one according to the same logic in (2.4) for sending the data packet.

Note that in order to meet various QoS requirements of users, the route selection logic in (2.4) can be modified accordingly to adapt to any QoS metric. To give an example, here we elaborate on the case where the route with maximum available bandwidth is of concern rather than the route with minimum delay when the network is dense enough, i.e., when $\exists$ route $j$: $1 / C_j < R$. In this case, any intermediate vehicle estimates the available channel capacity, denoted by $BW_i$ for vehicle $i$, based on a simple temporal approach. In this approach, every intermediate vehicle computes the percentage of channel idleness by dividing the total idle time of the channel during an observation interval by the length of the observation interval. The total idle time is the total time during which the vehicle neither transmits nor senses the channel busy. Improving the accuracy of available bandwidth estimation approaches for wireless ad hoc networks have been studied in many recent papers [49-51], but it is outside the scope of this thesis. Since the available bandwidth is a concave metric, it is only required that a route available bandwidth field be included in the RD and replaced by intermediate vehicles with smaller values. Thus, for every route $j$ the route available bandwidth, denoted by $RBW_j$, is obtained from,

$$RBW_j = \min_{\text{vehicle}\in\text{route}_j} (BW_i),$$

(2.5)
and the route selection logic in (2.4) is changed to,

\[
\text{route } k = \begin{cases} 
\arg \max_{j \in \mathcal{W}} (RBW_j) & \exists \text{route } j: 1/C_j < R \\
\arg \max_{j \in \mathcal{U}} (C_j) & \text{otherwise} 
\end{cases}
\] (2.6)

Note that the requesting vehicle may receive RR messages from multiple GWs, among which it uses the route selection logic to select the best one. Considering that the vehicle is constantly moving and the network conditions are continuously changing, it is very likely that the selected route is no longer the appropriate route. Therefore, the fields in the header of packets are updated every time a packet is forwarded between the requesting vehicle and the GW to determine if the current route is about to become invalid and a new route needs to be established.

2.3 Performance Limitation Scenarios and Solutions

2.3.1 Vehicle tracking mechanism for moving destinations

As mentioned earlier, the requesting vehicle places its location and velocity vector in the RD that it generates, and this information will also be included in the RR that the GW generates. By the time the RR (or any acknowledgement or data packet) is sent back to the requesting vehicle, there is the chance that it has moved far away from its initial location recorded in the RR (or the acknowledgement or data packet). Particularly, when the network is sparse and packets need to be partly carried by vehicles, the packet will return after a longer time, thereby increasing the chance of the target vehicle moving far away. Upon the arrival of the packet in the initial location of the requesting vehicle, if the vehicle responsible for forwarding the packet is not able to locate the requesting vehicle in its transmission range, it will forward the packet towards the velocity vector of the requesting vehicle recorded in the packet. Now, the problem occurs when the packet is forwarded to a junction, but the
requesting vehicle has already left that junction and there is no clue in which direction it has left which means the requesting vehicle cannot be tracked anymore. To resolve this issue, we propose the following vehicle tracking mechanism in the design of CMGR.

When the requesting vehicle arrives at a junction and turns onto one of the out-going roads, it attaches its new velocity vector to the next beacon message that it broadcasts. All the vehicles that hear this beacon message keep this information as long as they reside at that junction and rebroadcast it whenever they are about to leave the junction. Therefore, the information remains at the junction until either the session expires or no further vehicle resides at the junction. When the returning packet arrives at the junction, the corresponding vehicle responsible for forwarding the packet queries the new velocity vector of the requesting vehicle and by following the updated velocity vectors the packet is eventually delivered to the requesting vehicle.

To elaborate on the proposed protocol particularly when the vehicle tracking mechanism comes into play, an example is given in Fig. 2.1. Assume that the JS from the requesting vehicle (vehicle S) to the GW is calculated as $J_a, J_b, J_c, J_h$ according to the route selection logic. As the next step after route determination, vehicle S sends its data packet along the JS using a greedy forwarding algorithm in which each vehicle forwards the packet to the neighbour closest to the next junction in the JS. In Fig. 2.1 (a) the packet is forwarded according to this method all the way to vehicle P where a local maximum is detected and as a result, vehicle P starts carrying the packet until the GW comes into its transmission range in Fig. 2.1 (b). In the next step a reply packet, e.g., acknowledgement or data packet, is sent back to vehicle S along the same JS.
As depicted in Figs. 2.1 (c) and 2.1 (d), on the way to return two more local maxima take place in vehicles M and G, respectively. Note that although in Fig. 2.1 (c) vehicles J and K are within the transmission range of vehicle M, packet transmission is not possible due to the lack of line of sight. Finally the packet is delivered to vehicle A in \( J_a \). If the vehicle tracking mechanism is not in place, the packet will be dropped as vehicle S has already left the junction. However, when vehicle tracking mechanism is running on vehicles, vehicle S broadcasts its velocity vector upon leaving \( J_a \) which is heard by vehicle B. Also, vehicle B rebroadcasts the velocity vector as it leaves the junction which is heard by vehicle A. Hence, after receiving the packet, vehicle A sends it towards the velocity vector of vehicle S and the packet is finally delivered to vehicle S in Fig. 2.1 (e).
Fig. 2.1: An example illustrating CMGR and the vehicle tracking mechanism
2.3.2 Packet forwarding decision scheme at sparse junctions

In the process of forwarding a packet along its JS, when the network is sparse, a local maximum may occur which corresponds to the case where no next-hop vehicle towards the next intended junction in the JS is available. In this situation, as stated before, the vehicle will start carrying the packet to prevent packet dropping. Carrying the packet, however, is the only possible solution so long as the vehicle is not in the range of any junction. If the local maximum occurs at a junction, the option of packet forwarding to other out-going roads, which are not part of the JS but in which a vehicle is available at the decision-making instant, may yield routes that are more connected and have lower delays. Particularly, if we only rely on packet-carrying when a local maximum occurs at a junction, it is likely that the packet-carrying vehicle may turn onto an out-going road which is not even a part of the JS. Hence, employing an appropriate packet forwarding decision-making scheme at junctions is necessary. In this respect, a number of decision schemes have been proposed with the aim of delivering the packet through the fastest end-to-end route [38, 52, 53]. Note that in the packet forwarding decision scheme at junctions the objective is to forward the packet to the location of the target identified in the packet, whereas in the vehicle tracking mechanism the packet has already reached the location of the target, but cannot find the target.

In all the previous studies on packet forwarding in VANETs the roadmap of the desired region is mapped onto a weighted directed graph with the average delays of roads as the weights of edges. The average delays are obtained based on the average densities and velocities of the roads initially stored in the digital maps stored in the vehicles. The sum of weights are then used to obtain the packet forwarding priorities of the out-going roads, with the out-going road that yields the minimum delay end-to-end route assigned the highest
priority, the one with the second minimum delay assigned the second highest priority, and so on. We take the same prioritizing approach in our decision scheme. The problem with the previous studies is that due to highly dynamic network topology, at the arrival time of the packet at the junction the real average densities may be very different from those stored in the digital maps. To resolve this problem, we suggest the use of a scalable efficient data aggregation algorithm such as the one we proposed in [54, 55] for continuous update of average densities.

Both [52, 53] rely on the deployment of static nodes at all junctions which are responsible for storing arriving packets for a while hoping for the appearance of appropriate next-hop vehicles at the junction. However, we are looking for a fully ad hoc infrastructure-independent solution. To the best of our knowledge, VADD [38] is the only fully ad hoc packet forwarding decision scheme at junctions in the literature. The most critical issue with VADD among others is the fact that packet forwarding decisions are based on current moving directions of both forwarding and receiving vehicles which are subject to change shortly, i.e., as soon as the vehicles get past the center of the junction. Therefore, we propose our packet forwarding decision scheme at junctions which takes any promising forwarding chance into account.

We suggest that over its whole residing time at the junction, the vehicle currently responsible for forwarding the packet, forward the packet to a vehicle on the first priority out-going road if one is detected. Until a first priority vehicle is detected, the decision scheme when the packet-forwarding vehicle is before the center of the junction, and thus its next moving direction is not known, will be different from when it is after the center. Before the center of the junction, if no first priority vehicle is found, the packet-forwarding vehicle
will forward to any vehicle present in the second priority road. After forwarding it will not remove the packet from its buffer just in case any opportunity to forward the packet to a first priority vehicle comes up. A similar approach is adopted when a vehicle is available in a lower priority road. By using this approach we guarantee that every opportunity for fast delivery of the packet to the destination is taken into consideration. When the packet-forwarding vehicle has passed the center of the junction, it will either continue to carry the packet if it has turned onto the first priority road, or will look for vehicles on higher priority roads otherwise until it exits the junction.

A typical scenario of a local maximum at a junction is depicted in Fig. 2.2. Assume that vehicle A has detected a local maximum. Also, assume that the out-going roads towards the east, north and south are the first, second and third priorities, respectively. Since it does not find vehicle in the first priority, it looks for a vehicle in the next priority, i.e., the second priority, and as a result forwards the packet to vehicle B. After forwarding the packet it keeps it for forwarding to future arrivals in the first priority. Although multiple copies of the same packet are propagated in the network, since the packet duplications occur when the network is sparse, it will not make the network congested. The pseudo code of CMGR is presented in Fig. 2.3. The highlighted sections in Fig. 2.3 describe the vehicle tracking mechanism and packet forwarding decision scheme, respectively.
Fig. 2.2: A typical scenario using packet forwarding decision scheme

Run by all vehicles upon receiving a packet:
if received packet is an RD
  if current time – generation time < message lifetime
    attach its location, ρᵢ, rᵢ, BWᵢ to the RD
    rebroadcast the RD
  else
    drop the RD
  end if
else
  if next intended recipient is the target
    if the target is found
      forward the packet
    else
      run vehicle tracking mechanism:
        query the new velocity vector of the target
        forward the packet towards the target based on the updated velocity vector
    end if
  else if local maximum
    if on a straight road
      start carrying the packet
    else if at a junction
      run packet forwarding decision scheme:
        calculate packet forwarding priorities
        if any vehicle in the first priority is found
          forward the packet
        else if the vehicle is before the center of junction (has not yet reached the center of junction)
          if any vehicle in other priorities is found
            forward the packet
            keep the packet and continuously check for future arrivals at higher priorities
          end if
        else if the vehicle has passed the center of junction
          if the vehicle is in the first priority
            continue carrying the packet
          else if the vehicle finds any vehicle in a priority higher than
Run by GWs upon receiving RDs:
run route selection logic:
for all the RDs
TT\textsubscript{j} = reception time – generation time
C\textsubscript{j} is obtained from (2)
RBW\textsubscript{j} is obtained from (5)
end for
if minimum-delay route is required
route k is obtained from (4)
else if maximum-BW route is required
route k is obtained from (6)
end if
determine the JS
generate and send back the RR

Fig. 2.3: Pseudo code of CMGR

2.4 Performance Evaluations

2.4.1 Simulation settings

The street layout we used in the simulations is a grid layout derived from a real street map in the TIGER database [56] from US Census Bureau. For simulating the mobility of vehicles, we used the Simulation of Urban Mobility (SUMO) [57] microscopic street traffic simulation package. In SUMO, different types of vehicles can be defined. We used this feature to differentiate buses from cars which is of concern in A-STAR. For each line of buses, we defined the route the buses run along, the origin and destination of the line and the period with which buses are emitted in the line. Buses are removed when they reach the destination. Every street is assigned a range of speed according to the digital map, and a functionality that defines whether the street is a plain street, a source street or a sink street,
and a priority of usage. The priority of usage plays a role in computing the way-giving rules at junctions.

By selecting appropriate priorities of usage for different streets, we set the average vehicle densities of streets according to the initially stored values in the digital map. In order to obtain the initial values of vehicle densities in the digital maps of vehicles, we run the mobility simulator for a period of time without generating any packet until the vehicle densities are stabilized in all of the streets. We observed that after 2000 seconds, the error between the obtained vehicle densities and the initial densities became smaller than 5% of the initial values for all the streets in the simulation area. Thus, we start sending packets in the network after 2000 seconds and run the simulation for another 4000 seconds before we deliberately impose violations in vehicle densities. As mentioned earlier, the employed data aggregation algorithm helps update the digital maps when the topology of the network is highly dynamic. In order to factor in the improvements that the data aggregation algorithm makes over non-real-time digital maps, we deliberately simulated two accidents in the mobility simulator. The accidents occur 6000 seconds and 12000 seconds after the beginning of the simulation. The accidents result in increases in vehicle densities in the streets where they occur, and in all other streets in their proximity.

For any given number of vehicles, every vehicle is randomly injected onto one of the source streets. When it reaches a sink street, it is removed from the network and randomly regenerated on one of the source streets, and this procedure continues over the whole simulation period. All vehicular trips during the simulation time are saved in a log, which is then used to generate a mobility trace-file that is usable by Network Simulator 2 (NS-2) [58] with the help of Mobility model generator for Vehicular networks (MOVE) [59]. The
parameters we used in the mobility model and the wireless communications parameters are listed in Table 2.1. The simulation street layout is depicted in Fig. 2.4.

| Table 2.1: Mobility-related and wireless communications-related parameters |
|--------------------------------------------------|------------------|
| Simulation area                                  | 2440m * 2490m    |
| Average length of streets                        | 500m             |
| Number of vehicles                               | 100 – 400        |
| Average velocity                                 | 15 ~ 105 km/h    |
| Simulation time                                  | 20000 sec        |
| $R$ (Transmission range)                         | 250 m            |
| Radio model                                      | Two Ray Ground   |
| Traffic model                                    | CBR over 20 random vehicles |
| CBR rate                                         | 4 packets/sec.   |
| Data packet size                                 | 1 KB             |
| Beacon size                                      | 512 bits         |
| Beaconing frequency                              | 2 beacons/sec    |
| Data rate                                        | 1 Mbps           |
| MAC layer                                        | IEEE 802.11 DCF  |
| Max. Contention Window                           | 32               |
| Number of GWs                                    | 1–4              |

Fig. 2.4: The simulation street layout
2.4.2 Simulation results

The performance metrics that we consider in our evaluations are packet delivery ratio, packet delivery delay, and ratio of dropped data packets. The packet delivery ratio is the ratio of the packets the network layer delivers to its higher layer to the packets generated in the higher layer and passed onto the network layer for delivery, which accounts for packet dropping due to packet time-outs that take place due to disconnections and inability in finding target vehicles. On the other hand, the ratio of dropped data packets is the ratio of the total data packets dropped due to Medium Access Control (MAC) layer collisions and packet time-outs to the total data packets transmitted throughout the network either arriving from higher layers or being retransmitted due to collisions during the simulation time.

Another metric we could use for the evaluation of the routing protocols is the routing control overhead which any protocol imposes to the network. The overhead packets in CMGR include beaconing packets, RD and RR packets, or the packets generated for vehicle tracking. The number of beaconing packets is the same in all protocols. On the other hand, the number of RD and RR packets is exactly equal to the number of packets A-STAR and VADD generate in the location service mechanisms that they use to discover the destination locations, i.e., GWs in our case. Therefore, the only type of packets that make the overhead of CMGR higher than the other protocols are those generated in the vehicle tracking mechanism. However, since these packets, unlike beaconing packets, are not generated periodically, and since these packets, contrary to RD and RR packets which are generated once or several times per session, are generated much less frequently and mostly when the network is sparse, the portion of the overhead traffic contributed by this type of packets is relatively low and can be neglected compared to beaconing and RD and RR packets. As the
control overhead is approximately similar in all the routing protocols, we do not present any numerical results for this performance metric.

Different end-user multimedia categories for a variety of multimedia services were investigated in [57] and it turned out that the maximum allowable one-way transmission delays for a relatively large number of multimedia services is either 10 seconds or 1 minute. Thus, we set the allowable message lifetime in our simulations accordingly. Furthermore, we run the simulations with different number of GWs. As stated before, since A-STAR and VADD are the only existing routing protocols that take connectivity into account, we only simulated these two protocols for performance comparisons. Since A-STAR prioritizes the streets with more bus lines in selecting routes, most of the data traffic will be pushed towards those streets increasing the chance of data traffic congestions on them. Besides, the number of bus lines on streets is not an accurate metric for assessing their connectivity.

In the first round of simulations only one GW exists in the network and is placed at the bottom rightmost junction of the network. The packet delivery ratios and the packet delivery delays for maximum allowable one-way delays of 1 minute and 10 seconds with 95% confidence intervals are depicted in Figs. 2.5 and 2.6, respectively. In the study of routing protocols, the common trend to ensure the reliability of the simulation results obtained is to first find the interval, specified as the window between low and high error bars, within which 95% of observations fall and then take the average of all the observations within that interval. Every point in Figs. 2.5 and 2.6 is obtained according to the same procedure. Note that the main reason the performance metrics are depicted in terms of the number of vehicles, which is equivalent to the density of vehicles in the network, is that one of the main challenges that the proposed protocol is addressing is adapting to both sparse and dense situations. In other
words, CMGR is designed to perform route selection based on connectivity-awareness in sparse situations, whereas it performs route selection based on lowest trip times when the network is dense and congestions should be avoided.

When a local maximum occurs, A-STAR computes another route as a recovery route. However, since vehicles do not carry packets in A-STAR, a packet is dropped in case no recovery route is found or after a limited number of recoveries. This best explains its low packet delivery delays and low delivery ratios compared to VADD and CMGR. Higher packet delivery ratios of CMGR in lower vehicle densities in Figs. 2.5 (a) and 2.6 (a), compared to those of VADD and A-STAR demonstrate the effectiveness of the vehicle tracking mechanism and packet forwarding decision scheme in CMGR.

![Graphs](image_url)

**Fig. 2.5:** (a) Packet delivery ratio for maximum delay of 1min. and 1 GW, and (b) Packet delivery delay for maximum delay of 1min. and 1 GW

![Graphs](image_url)

**Fig. 2.6:** (a) Packet delivery ratio for maximum delay of 10sec. and 1 GW, and (b) Packet delivery delay ratio for maximum delay of 10sec. and 1 GW
As evident from Figs. 2.5 (a) and 2.6 (a), the superior performance of CMGR in lower vehicle densities is more noticeable when the maximum allowable delay is 1 minute. This can be well-justified by taking a closer look at the vehicle tracking mechanism. The improvement that the vehicle tracking mechanism makes becomes less considerable as the maximum delay decreases, because for lower maximum allowable delays the corresponding packet delivery delays are smaller, and in such a short time the route-requesting vehicles will not get the chance to go far away from their initial locations. Also higher delivery delays of CMGR are a direct result of its higher delivery ratios. Note that despite its higher average delivery delays compared to A-STAR and VADD, CMGR delivers the same packet faster than both A-STAR and VADD. However, the higher average delivery delays in CMGR is due to the fact that the packets that cannot be delivered by A-STAR or VADD but are delivered by CMGR mostly undergo longer delays and therefore increase the average packet delivery delays. As a result, higher average delivery delays of CMGR at lower vehicle densities compared to A-STAR and VADD with respect to its higher average delivery ratios do not imply a tradeoff between average delay and average delivery ratio.

The ratios of dropped data packets are given for maximum allowable one-way delays of 1 minute and 10 seconds in Figs. 2.7 and 2.8, respectively. As observed in the figures the increase in the number of vehicles in the network affects the performance of the network in two ways. On one hand, it increases the connectivity which in turn reduces the number of packets dropped due to local maximums. On the other hand, the presence of more vehicles in the network makes it more congested as a result of more data and beacon packet transmissions, thereby increasing the chance of MAC collisions which in turn increases the number of dropped packets.
In the second round of simulations, we investigate how the deployment of more GWs in the network affects the results. We place the GWs at the farthest possible distance from each other to make sure the vehicles in the network benefit from the added GWs most. For instance, in the case of two GWs, we place one at the top leftmost junction and the other one at the bottom rightmost junction of the network. As it can be extrapolated from Fig. 2.5 (a), the packet delivery ratios of CMGR with maximum delay of 1 minute are very close to one for all vehicle densities when the number of GWs in the network is greater than one. Therefore, for maximum delay of 1 minute we only present the graph of packet delivery delays (Fig. 2.9). The packet delivery ratios and delivery delays of CMGR with maximum delay of 10 s for different number of GWs are depicted in Figs. 2.10 (a) and 2.10 (b), respectively. The results were obtained with 95% confidence intervals.

As it can be observed from Figs. 2.9 and 2.10 (a), the performance improvement obtained by using more GWs becomes less considerable when average vehicle density increases. Based on these results, for any required packet delivery ratio and delivery delay, the number of GWs that should be installed in any given area of the network could be determined relative to the average vehicle density in that area.
Fig. 2.9: Packet delivery delay for maximum delay of 1min. and different number of GWs

Fig. 2.10: (a) Packet delivery ratio for maximum delay of 10sec. and different number of GWs, and (b) Packet delivery delay for maximum delay of 10sec. and different number of GWs

2.5 Summary

In this chapter, we have presented a connectivity-aware minimum-delay routing protocol (CMGR) for VANETs that employs a novel route selection logic adaptable to the continuously changing density of vehicles in the network. In order to deal with network disconnections in sparse situations, the routes with higher vehicle densities should be prioritized to make the chance of packet delivery is maximized. However, prioritizing the routes with higher densities in dense situations increases the likelihood of packet traffic congestion. As a result, in dense networks, among the routes with enough connectivity level, less congested routes should be favored. For this purpose in the proposed routing protocol, among all the routes with enough connectivity, the routes with minimum delays or the routes
with maximum available bandwidths have been selected. We have concluded that the flexibility of the routing protocol, i.e., the adaptability of the routing protocol mechanism to the density of vehicles in the network, is a key design concept and is essential to guarantee the efficiency of the routing protocol.

Also, due to the high levels of mobility in the network, target vehicles are very likely to leave their initial locations, i.e., the locations known to the other nodes that are currently maintaining an active session with them. Hence, many packets addressed to these target vehicles are sent to their obsolete locations and are dropped. In order to deal with this very common issue, a target tracking mechanism has been included in our proposed protocol that shows a promising performance improvement which makes us believe that the a target tracking mechanism is a key element in the design of any successful routing protocol for vehicular environments.

Also, the other important aspect in the design of a routing protocol for vehicular networks is the forwarding algorithm that is used at intersections, particularly when the intersection is so sparse that no next-hop vehicle can be found on the intended out-going road. In other words, making wise decisions upon packet forwarding at intersection by taking every promising forwarding opportunity into account in an efficient way in terms of redundant packet traffic is a significant design challenge. For this purpose, in the proposed routing protocols a decision scheme has been introduced for forwarding packets in sparse junctions which aims to minimize end-to-end delays by taking every promising forwarding opportunity into account. At the end, CMGR has been compared with two plausible geographic connectivity-aware routing protocols for VANETs, A-STAR and VADD. The obtained results show that CMGR outperforms A-STAR and VADD in terms of both packet
delivery ratio and ratio of dropped data packets. For example, under the specific conditions considered in the simulations, when the maximum allowable one-way transmission delay is 1 min and one gateway is deployed in the network, the packet delivery ratio of CMGR is approximately 25% better than VADD and A-STAR for high vehicle densities and goes up to 900% better for low vehicle densities. In addition to the simulation results, the main challenges that were addressed by CMGR and the corresponding features employed in CMGR to address them are listed in Table 2.2 to highlight the main differences between CMGR and previous routing protocols.

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simultaneous adaptability to both sparse and dense situations</td>
<td>CMGR differentiates sparse situations from connected situations: connectivity-aware route selection to favor routes with higher connectivity in sparse situations and QoS-aware route selection to avoid congested routes in dense situations</td>
</tr>
<tr>
<td>Metrics to assess connectivity should be real-time as opposed to offline databases</td>
<td>In CMGR real-time connectivity information calculated on-the-fly upon route discovery phase</td>
</tr>
<tr>
<td>Situations where source vehicle has turned a corner</td>
<td>A vehicle tracking mechanism developed in CMGR to keep velocity vector of source in intersection until route reply comes back</td>
</tr>
<tr>
<td>Reliable packet forwarding at sparse intersections</td>
<td>A packet forwarding decision scheme for intersections developed in CMGR to guarantee that every opportunity for fast delivery of packets is considered</td>
</tr>
</tbody>
</table>
Chapter 3: Optimal Distributed Vertical Handoff Strategies in Vehicular Heterogeneous Networks

3.1 Introduction

The extensive development of wireless communication systems has resulted in the availability of several access technologies at any geographic area, such as cellular networks, wireless WLANs and wireless broadband networks. This heterogeneous wireless environment can help meet the diverse QoS demands of end users. Vehicular networks will particularly benefit from such a rich set of connectivity options as various types of vehicular infotainment applications are expected to be simultaneously accessed by vehicular users and the use of a single wireless access network may not suit the needs of all vehicular applications. Therefore, developing VHNs, comprised of different wireless access technologies which may be solely dedicated to vehicular communications or else are part of a wider public network, is a priority in the near future. Some further beneficial consequences of integrating different wireless access technologies include the more efficient usage of the resources of such heterogeneous networks, extending the coverage of service availability and offering a range of connectivity alternatives, in terms of QoS support, coverage areas and service costs. In this chapter we consider a VHN consisting of a wide-area cellular network for global access complemented by WLANs with fixed or random inter-distances between access points placed along roadsides as shown in Fig. 3.1. The cellular network provides coverage to areas not covered by the access points.

Even though VHO maintains active connections of nodes while moving at high speeds by providing the possibility to switch from one access technology to another, it also enables novel types of applications to be developed, especially in a vehicular network. The VHO
decision, i.e., selecting the most appropriate access technology among available alternatives for a particular application can depend on both user preferences, in terms of perceived QoS, cost and/or battery lifetime, among other parameters, and network preferences such as load balancing, interference avoidance and revenue maximization.

Fig. 3.1: The reference model

In this respect, many VHO decision-making algorithms have been proposed in the literature [7-11, 39-43]. However, extension and adaption of such solutions in the contexts of VHNs has received relatively less attention. In [39] users select the wireless access technology with the highest bandwidth as the most appropriate option. The authors of [40] formulated a multi-objective optimization problem and proposed a heuristic with the aim of selecting a network connection for transferring each component of data. In [41, 42] each user provides the network with up to ten different inputs to assist the network in making a VHO
decision based on its specific preferences. Half of the input values are weights describing the importance of VHO decision-making parameters including cost, security, power, network conditions and network performance to the user. The rest of the inputs are threshold values specifying allowable range of the VHO decision parameters. In [7], available access networks are first characterized as acceptable if they satisfy the minimum cut-off criteria, and unacceptable otherwise, by using a non-compensatory Multi-Attribute Decision-Making (MADM) algorithm. Then a compensatory MADM algorithm is used to calculate the rankings of the acceptable networks based on their costs, available bandwidths, allowed bandwidths, utilizations, delays, jitters, and packet losses. The authors of [8] proposed a VHO decision algorithm which focuses on the joint optimization of total battery lifetime of the network, and fair distribution of traffic load at attachment points which improves the overall QoS by avoiding congestions.

When it comes to VHNs, which have highly dynamic network topologies and highly variable environment conditions due to the inherent characteristics of high mobility vehicular communications, the VHO decision-making algorithms mentioned above might be inefficient and ineffective. This inefficiency stems from the fact that in the design of the VHO decision-making algorithms the mobility models of users including their movement trajectories and their velocities are often neglected. To emphasize the role of mobility pattern awareness, note that when traveling at high speeds, it is more likely that one user travels through several access technologies in a short span of time. Therefore, when legacy VHO decision-making algorithms are being used, it is highly probable that handing off from a wide-coverage network to a newly emerged local-coverage network may be followed by another VHO back to the original network immediately afterwards resulting in too many VHOs. Since the
procedure of a VHO involves a set of signalling functions and consequently imposes both VHO processing loads and signalling overhead to the network, unnecessary VHOs should be discouraged. Overloading the network with signalling traffic in turn causes additional costs and longer transfer times incurred by delays of reconnecting the user to the new network. Another aspect of the mobility models that has been neglected in most previous studies is the fact that the movements of vehicles are confined by roadways, so that the directions of movements are highly constrained and only the network coverage along these directions of movements is of interest.

Very few studies have exploited the knowledge of mobility patterns of users in a VHO decision-making mechanism for VHNs. In [43], the authors utilized a centralized Location Service Server (LSS) to which vehicles report their current positions and consequently receive the information of available access networks in their vicinity. Further, a utility function is used to determine the satisfaction of users in accessing the available networks, which is fed back to LSS. Finally, the optimal hand-off decisions are periodically calculated in LSS at the beginning of discrete time intervals based on nodes movement predictions and are reported back to the users. To the best of our knowledge, the decision-making mechanism that we present in this chapter is the first work in which vehicles continuously select optimal access networks, based on their mobility patterns, in a distributed event-based manner.

A final issue of interest in analyzing VHO is the signalling traffic associated with this process. The VHO signalling traffic can be transmitted either on an existing access technology or via a dedicated wireless signalling system as in [61]. The authors of [61] suggest the use of a two-way paging system for signalling negotiations of VHO decisions. Since we assume that all areas of the network are covered by the cellular networks, in our
study we use cellular systems for signalling transmissions so that its location and mobility management functionalities can be shared with other access networks without any extra deployment costs. For this purpose, a combination of tight and loose coupling can be used. Note that in loose coupling different access networks are independent and are connected to each other through the Internet. However, in tight coupling the networks with smaller coverage areas attach to the network with larger coverage area in the same manner a radio access network attaches to the core network, and are dependent on the larger network in that all of their signalling functionalities and data transfers are handled by the larger network. By using a combination of tight and loose coupling the cellular core network considers WLANs as part of its access network to enable signalling traffic transfer through the cellular network. Different architectures for tight and loose couplings and their pros and cons are studied in more detail in [12, 76, 93].

Main Contributions: In this chapter we develop an optimal, event-activated, and thus continuous-time, VHO decision-making algorithm, which is based on the mobility profiles of users including their velocities, and further takes the preferences of users in terms of costs or transfer times into account. As opposed to most existing solutions, the proposed approach is deterministic and is fully distributed in the sense that vehicular users will make the VHO decisions rather than network entities. We address handoff decision-making in a comprehensive set of system models, pertinent to infrastructure-based access technologies on the roadside, V2R communications, as well as in scenarios where both V2R and ad hoc communications between vehicles, V2V communications, are feasible. Furthermore, we obtain the optimal VHO decision-making mechanism when planned WLANs for vehicular communications exist at certain areas, where the locations of access points are known a priori.
by the vehicular users, and extend the analysis to the case of open WLAN access points that are randomly located. To the best of our knowledge, our contribution is the first rigorous study of VHO decision-making, addressing such a comprehensive set of scenarios.

The rest of this chapter is organized as follow. The details of the problem formulation are elaborated in Section 3.2. Optimal VHO decisions with fixed inter-distance access points are studied in Section 3.3, followed by an analysis for random inter-distance access points in Section 3.4. Both V2R and V2V communications are considered in Section 3.5. Numerical results are presented in Section 3.6. Section 3.7 summarizes the chapter.

3.2 Problem Formulation

We focus on a scenario where cellular networks and WLANs interwork to form a VHN. However, our proposed solution can be readily extended to any arbitrary set of access technologies. Although recently there have been a lot of studies on the integration of various wireless access technologies, our motivations for this choice include the widespread availability of dual-mode terminals, the relatively long history of coexistence of these technologies, and the complementary characteristics of cellular and WLANs, whereby the cellular system provides a larger coverage area with a relatively lower data rate at a higher cost whereas a WLAN can offer a relatively higher data rate in a shorter range with a lower cost. In the presence of both WLAN and cellular network, the vehicle has the option of picking either of them for communications, whereas upon leaving the WLAN coverage the cellular network will be the only alternative, if V2V communication is not feasible. We will extend our analysis to consider both V2R and V2V communications in Section 3.5.

Every vehicle that wishes to access the VHN should establish a connection to an attachment point, i.e., a BS on the cellular network or an AP of the WLAN. Since we assume
that the cellular network provides global coverage, the vehicle can always find a BS with a strong enough level of Received Signal Strength (RSS). Fig. 3.1 depicts a reference model for the architecture described.

If the RSS of an AP starts decaying rapidly and falls below a threshold for a specific period of time, the vehicle will initiate the VHO to the cellular network. All BSs and APs periodically broadcast advertisement messages announcing their availability and their prices of data transfer, denoted by \( c_w \) and \( c_c \) ($\) as the costs of sending one bit in WLAN and cellular network, respectively. In practice, service providers price their services based on business considerations, among which use of capacity is but one of the factors. However, we approach the problem at hand from a system design perspective to determine the minimal time or cost of communication for the vehicular users. Deriving such lower-bound access time or cost limits will facilitate profitable pricing strategies for service providers. As both WLAN and cellular technologies support periodic network identification signalling, addition of access cost and access rate information will have a negligible effect on the system overhead. Note that as we assume that cellular network is responsible for transferring the signalling traffic, the cost of sending a signalling bit will also be \( c_c \). Clearly, both access networks have limited capacities and therefore their costs are dependent on the available resources at the time instant they broadcast their advertisement messages. Therefore, one objective of the pricing mechanism can be adaptive load balancing; however, determination of optimal pricing strategies is out of the scope of this chapter. Note that BSs and APs could be under the ownership and control of different service providers and consequently follow different policies for the determination of their costs. A glossary of all variables and their definitions is given in Table 3.1.
Table 3.1: Variables and their definitions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( c_w )</td>
<td>Cost of sending a bit in WLAN</td>
</tr>
<tr>
<td>( c_c )</td>
<td>Cost of sending a bit in cellular network</td>
</tr>
<tr>
<td>( c_1 )</td>
<td>Transmission cost of only cellular networks</td>
</tr>
<tr>
<td>( c_2 )</td>
<td>Transmission cost of WLANs + cellular</td>
</tr>
<tr>
<td>( c_3 )</td>
<td>Transmission cost of WLANs + ad hoc</td>
</tr>
<tr>
<td>( c_4 )</td>
<td>Transmission cost of WLANs + cellular + ad hoc</td>
</tr>
<tr>
<td>( r_w )</td>
<td>Data rate in WLAN</td>
</tr>
<tr>
<td>( r_c )</td>
<td>Data rate in cellular network</td>
</tr>
<tr>
<td>( W )</td>
<td>Coverage area of each WLAN AP</td>
</tr>
<tr>
<td>( A )</td>
<td>Distance between two consecutive AP coverage areas</td>
</tr>
<tr>
<td>( A' )</td>
<td>Distance over which ad hoc delay is tolerable</td>
</tr>
<tr>
<td>( b_t )</td>
<td>Data bits required to be sent</td>
</tr>
<tr>
<td>( b_{VHO} )</td>
<td>Signaling bits required for a VHO</td>
</tr>
<tr>
<td>( N_w )</td>
<td>Number of WLANs needed for transmitting all ( b_t ) bits</td>
</tr>
<tr>
<td>( T_w )</td>
<td>Total transmission time when WLANs + cellular</td>
</tr>
<tr>
<td>( T_c )</td>
<td>Total transmission time when only cellular networks</td>
</tr>
<tr>
<td>( T_u )</td>
<td>Usage time</td>
</tr>
<tr>
<td>( T_{w+AH} )</td>
<td>Total transmission time when WLANs + ad hoc</td>
</tr>
<tr>
<td>( T_{c+w+AH} )</td>
<td>Total transmission time when WLANs + cellular + ad hoc</td>
</tr>
<tr>
<td>( th )</td>
<td>Pre-determined threshold adjusting the accuracy of VHO</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>Arrival rate of WLAN coverage areas</td>
</tr>
<tr>
<td>( t_i )</td>
<td>Inter-arrival times with lengths smaller than ( W )</td>
</tr>
<tr>
<td>( D_i )</td>
<td>Inter-arrival times with lengths greater than ( W )</td>
</tr>
<tr>
<td>( d_{AH} )</td>
<td>Ad hoc communication delay</td>
</tr>
<tr>
<td>( \rho )</td>
<td>Average vehicle density</td>
</tr>
<tr>
<td>( v_h )</td>
<td>Average vehicle velocity</td>
</tr>
<tr>
<td>( d_{hop} )</td>
<td>Average delay per hop</td>
</tr>
</tbody>
</table>

3.3 VHO Decision-making Algorithm with Fixed AP Inter-distances

Since VHO decision-making requires a relatively high level of transmission power and processing power capabilities, in most studies the decision-making procedure is carried out in a fixed network processing centre, i.e., in a centralized manner. However, for vehicular networks such power constraints can be relaxed and consequently in this chapter we propose a distributed VHO decision-making algorithm which removes the need for deploying a data-processing and decision-making centre in the core network and the packet traffic between the centre and nodes. In this distributed setting, every vehicle, based on the information initially
loaded in its database and the inputs repeatedly updated by the network, will make VHO decisions. Provided that at the time instant in which the VHO decision is being made, \( b \) data bits are required to be transmitted and given that both the WLAN and the cellular network are available to the vehicle, the VHO decision-making algorithm should decide which one to access depending on the user’s preferences. These preferences can include minimization of the transmission cost or alternatively the transmission time. Note that even when the WLAN is prioritized over the cellular network, using the cellular network in areas that are not covered by the WLAN is inevitable and this is why this case is referred to as \textit{WLAN plus cellular} throughout the chapter. In the case that the vehicle’s first priority is to minimize the transmission cost, the decision-making algorithm should calculate the cost of transmission for both accessing the cellular network, i.e. \( c_1 \) ($), or else the WLAN (WLAN plus cellular), i.e., \( c_2 \) ($), and hence selects the access network with the lowest cost. In subsection 3.3.1 the case of cost minimization VHO strategy is discussed, which is followed by transmission-time minimization analysis in subsection 3.3.2.

\textbf{3.3.1 Cost-minimization approach}

We begin with a set of simplifying assumptions to define the cost functions, but as we move forward, we will relax the assumptions and improve the functions accordingly to make our formulations more realistic. We assume that all vehicles are equipped with both WLAN and cellular interfaces and the allowable data rates announced by the WLAN and cellular at the time of decision-making are \( r_w \) and \( r_c \) (bps), respectively. Further, we consider a highway vehicular communication scenario, where without loss of generality we can confine our analysis in spatial domain to the direction of moving vehicles, and therefore in the rest of chapter we address the space in (m) rather than (m\(^2\)). The coverage area of each AP is
assumed fixed and equal to $W$ (m) implying a planned network. Furthermore, the APs are all equidistant and the distance between the coverage areas of two consecutive APs, which is covered by cellular network, is equal to $A > 0$ (m). The scenario of a vehicle driving along a highway including the defined parameters is depicted in Fig. 3.2. Although we derive all formulas based on the highway scenario, the extension to an urban area with a number of intersecting streets is straightforward.

Assuming that the velocity of the vehicle at the decision-making time is $v$ (m/s), the access costs $c_1$ and $c_2$ are obtained as

$$c_1 = b_t c_c,$$  \hspace{1cm} (3.1)

$$c_2 = N_w \left( \frac{W}{v} \right) r_w c_w + (T_w - N_w \left( \frac{W}{v} \right)) r_c c_c,$$  \hspace{1cm} (3.2)

where $N_w$ is the number of AP coverage areas the vehicle drives across before all the $b_t$ bits are transmitted and $T_w$ is the total time needed for transmission in the case where the WLAN is prioritized over the cellular network (WLAN plus cellular). Note that $b_t$ does not include the VHO signalling bits. The first term on the right hand side of (3.2) is the cost of transmitting/receiving data-only bits over the WLAN. The second term in $c_2$ is the incurred cost of transmitting data, and signalling bits, over the cellular network. Therefore, given that...
$b_{VHO}$ is the number of signalling bits needed for a VHO, $N_w$ is the maximum integer value so that

$$N_w \frac{W}{v} r_v + (N_w - 1) \frac{A}{v} c_v - 2N_w b_{VHO} \leq b_t,$$

and $T_w$ is given by

$$T_w = \left[ x + N_w W + (N_w - 1) A \right] / v,$$

where $x$ is the distance depicted in Fig. 3.2 and can be obtained from the equation below

$$b_t - \left[ N_w \frac{W}{v} r_v + (N_w - 1) \frac{A}{v} c_v - 2N_w b_{VHO} \right] = \frac{x}{v} c_v.$$

Note that to derive (3.3), we assume that even if the user is initially within the range of an AP, before it starts to access any network, the access initialization signalling, which is taken care of by the cellular network, is comparable to VHO signalling overhead, and thus irrespective of the initial location of vehicular user $2N_w b_{VHO}$ signalling bits are required. When the velocity goes up, the number of required VHOs increases which in turn may cause an unreasonable amount of VHO signaling traffic compared to the data traffic and consequently a higher cost of data transmission. By computing the costs and selecting the access technology with the minimum cost the VHO decisions are adapted to the velocity of the vehicle accordingly.

### 3.3.2 Transmission-time minimization approach

Under some circumstances, the vehicular user’s preference could be accessing the technology with the highest QoS metrics. Among various QoS metrics, data rate can be considered of significant importance. Although the data rates offered by the WLAN and the cellular network are out of the user’s control, by choosing appropriate access network at any point, the total transmission time of the data bits can be minimized. Therefore, using the
same approach discussed in calculating the costs, the vehicle can calculate the transmission times $T_c$ and $T_w$ (s) which are the total times needed for the transmission when only the cellular network and WLAN plus cellular are used, respectively.

\[
T_c = \frac{b_i}{r_c}
\]

(3.6)

and $T_w$ is given by (3.4). The vehicle selects the network with the minimum transmission time. It is worth mentioning that the provisioned decision-making process is event-based, whereby the aforementioned parameters will be recomputed upon observation of any significant change in the network such as advertisement of new costs or new allowable data rates, availability of new alternative access network, or following a considerable change in the velocity or direction of vehicle.

3.4 VHO Decision-making Algorithm with Statistical AP Inter-distances

IEEE 802.11-based WLANs have been extensively deployed in home and offices around the world. Since upstream access links of these networks are often idle, they could potentially be used for providing service to vehicles. The possibility of using such an unplanned set of open WLANs in terms of security, viability, and deployment to offer services to end users moving at vehicular speeds has already been studied in the literature [62, 63]. Since open APs are independently deployed along roadsides and no vehicles have prior information about their placements, we assume that the distances between consecutive APs follow negative exponential distribution. In other words, when a vehicle is moving with a fixed velocity, APs will show up in its transmission range according to a Poisson arrival. This assumption is also aligned with real-life measurements of open WLAN coverage for vehicular networks such as [62]. So, in this section we extend our analysis in Section 3.3 to
the scenario where the distances between various APs change randomly according to a Poisson process.

In case of an overlap where a new WLAN shows up before the vehicle exits the previous one, we assume that the vehicle continues its connection with the first AP and handoffs to the new AP when it is no longer covered by the first AP. Therefore, we can assume that the area covered by WLAN APs in this scenario is the sum of the areas covered by each AP, minus the overlapping area. Assuming that the coverage area of each AP is still fixed, and equals \( W \) as in previous section, the areas covered by the overlapping APs will be a random variable with a general distribution. A typical network topology in this new scenario is depicted in Fig. 3.3.

In this case, \( N_w \) is the maximum integer value satisfying

\[
\frac{W_1}{v} r_w + \frac{A_1}{v} r_c + \frac{W_2}{v} r_w + \frac{A_2}{v} r_c + \ldots + \frac{A_{N_w-1}}{v} r_c + \frac{W_{N_w}}{v} r_w - 2N_w b_{VHO} \leq b_t. \tag{3.7}
\]

Note that both \( W_s \) and \( A_s \) are stochastic variables. Hence, \( N_w \) will be the maximum integer value such that

\[
P(M \leq b_t) \geq \text{th}, \tag{3.8}
\]

which is equivalent to

\[
P(M \geq b_t) \leq 1 - \text{th}, \tag{3.9}
\]

where \( P(x) \) denotes the probability of event \( x \) and \( \text{th} \) is a pre-determined threshold adjusting the accuracy of VHO decisions. Since \( M \) is a non-negative random variable and \( b_t \) is greater than zero, according to Markov’s inequality [64] which provides a tight bound we have

\[
P(M \geq b_t) \leq \frac{E(M)}{b_t}, \tag{3.10}
\]
where $E(x)$ denotes the expected value of $x$. Therefore, if we obtain the maximum integer value for $N_w$ such that $E(M) / bt \leq 1 - th$ holds, (3.8) will hold as well. The idea is that for a given $N_w$, we compute $E(M)$ and compare it with $bt(1 - th)$. Then, $N_w$ is the maximum integer value for which the above inequality holds. Now, we explain how $E(M)$ is computed for a given $N_w$. For any given $N_w$ we have,

$$E(M) = N_w \frac{E(W_i)}{V} r_w + (N_w - 1) \frac{E(A_i)}{V} r_c - 2N_w b_{\text{FHQ}},$$

(3.11)

where $E(W_i)$ is as follows.

$$E(W_i) = E(W_i, 0 \text{ AP showup in } W) + E(W_i, \text{ at least 1 AP showup in } W)$$

$$= E(W_i, 0 \text{ AP showup in } W) + E(W_i, 1 \text{ AP showup in } W \text{ and } 0 \text{ AP showup in } W \text{ after the first showup})$$

$$+ E(W_i, 1 \text{ AP showup in } W \text{ and at least 1 AP showup in } W \text{ after the first showup})$$

$$= ...$$

$$= W P(0, W) + [W + E(y_i)] P(0, W) P(1, W) + [W + 2E(y_i)] P(0, W) [P(1, W)]^2 + ...$$

(3.12)
where given a Poisson arrivals of APs with rate $\lambda$ [65] the probability of $k$ arrivals in time $t$
eq t
$$P(k,t) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}.$$  

(3.13)

Furthermore, $y_i$s as shown in Fig. 3.3 are the inter-distances of APs in the areas covered by the WLANs, which have a negative exponential distribution with average $1/\lambda$. However, since $y_i$s are all smaller than $W$, their expected value $E(y_i)$ will be

$$E(y_i) = E(y | y < W) = \frac{\int_0^W y f(y) dy}{P(y < W)} = \frac{(-W e^{-\lambda W} + \frac{1}{\lambda} - \frac{1}{\lambda} e^{-\lambda W})}{(1 - e^{-\lambda W})} = f(y) = \lambda e^{-\lambda y}.$$  

(3.14)

Therefore, $E(W_i)$ in (3.12) can be simplified as

$$E(W_i) = \frac{W e^{-\lambda W} (1 - \lambda W e^{-\lambda W} - e^{-2\lambda W})}{(1 - e^{-\lambda W})^2 (1 - \lambda W e^{-\lambda W})^2}.$$  

(3.15)

Also, $E(A_i)$ in (3.11) is given by

$$E(A_i) = E(D_i - W) = E(D_i) - W,$$  

(3.16)

where $D_i$s as shown in Fig. 3.3 are the inter-distances of APs when no AP has showed up for at least $W$ and consequently $D_i$s are greater than $W$. Knowing that the arrivals of APs are Poisson, for $E(D_i)$ we have

$$E(A_i) = E(y | y > W) = \frac{\int_0^\infty y f(y) dy}{P(y > W)} = \frac{W}{W} = W + \frac{1}{\lambda} - W = \frac{1}{\lambda}; f(y) = \lambda e^{-\lambda y}.$$  

(3.17)

In the following subsections, we address the VHO decision-making problem with cost minimization as well as transmission time minimization goals.
3.4.1 Cost-minimization approach

The computed $N_w$ is used to obtain costs and transmission times required in the VHO decision-making algorithm. When the objective is selecting the access network with the minimum cost, and given the computed $N_w$, the expected value of $c_2$ which was used in decision-making can be obtained in a similar way to (3.2) as

$$E(c_2) = N_w \frac{E(W_i)}{v} r_w c_w + [T_w - N_w \frac{E(W_i)}{v}] r_c c_c. \quad (3.18)$$

Similarly, from (3.4), we have

$$T_w = [x + N_w E(W_i) + (N_w - 1)E(A_i)]/v, \quad (3.19)$$

where $x$ is obtained in the same way as of (3.5), i.e.,

$$b_i - [N_w \frac{E(W_i)}{v} r_w + (N_w - 1) \frac{E(A_i)}{v} r_c - 2N_w b_{HO}] = \frac{x}{v} r_c. \quad (3.20)$$

3.4.2 Transmission-time minimization approach

When the vehicle's preference is the access network resulting in the minimum transmission time, the decision-making algorithm should compare $T_c$, which is calculated according to (3.6), with $T_w$ as given by (3.19) to select the optimal access strategy. In Section 3.6 we will benchmark the performance of the proposed VHO decision-making algorithm in case where AP inter-distances are statistically distributed with the fixed AP inter-distances case.

3.5 VHO Decision-making with Enabled V2V Mode

We now generalize our proposed VHO decision-making algorithm to include the scenarios where multi-hop V2V communications are also allowed between vehicles, in addition to V2R communications, in the architecture of the VHN. Using intermediate vehicles to relay data to APs via ad hoc communications can alleviate the need for accessing
the costlier cellular network when APs are out of range. There are few studies in the literature that include WLAN ad hoc mode in their architecture. One example is [8] where the authors have devised a route-selection algorithm to forward data to attachment points. However, contrary to our algorithm, which considers multi-hop ad hoc communication as a data transmission alternative in addition to cellular or WLAN plus cellular, the previous work employs ad hoc networking only as a means for forwarding data to the attachment points which have been pre-selected.

Since the delays of V2V communications consists of multi-hop relaying delays and the delays experienced when no next hop vehicle is found and the data is being carried, these delays are generally much longer than the communication delays in the core network, which has a high-speed wired backbone. Thus, the delays imposed by ad hoc networking should be well investigated for delay-sensitive applications in order to ensure meeting their maximum allowable delay requirements. To this end, we propose an ad hoc delay calculation scheme in the following. Note that what we mean by continuous event-based VHO decision-making strategies is that the vehicular user initiates decision-making process as soon as the level of received signal from its existing access network degrades for a certain period of time, or any significant changes in the network setting is detected. Therefore, it is logical to assume that VHO delay will not contribute towards the ad hoc delay calculated in this section.

3.5.1 Ad hoc networking delay analysis

We assume that vehicles are equipped with GPS receivers and thereby they have accurate knowledge of their geographical positions. Also, they periodically broadcast beacon messages reporting their positions to their neighbours and based on these beacons they maintain an updated neighbour list in their look-up tables. Similar to the previous section, we
start with the case where the APs’ inter-distances are fixed and then extend the results to the scenario where APs show up randomly with an identical independent distribution.

A typical scenario when the distances between consecutive APs are fixed and equal to \( d = A + W \) is depicted in Fig. 3.4. In this section we will take the worst-case ad hoc communication delay into account. Alternatively, the expected (average) ad hoc delay could be calculated. However, since our approach in this work is to develop a deterministic VHO decision-making policy, if the expected delay is considered instead of the worst-case delay in the decision-making, it might result in outage in connectivity due to instances of excessive delay over the average delay in the network. The worst-case delay profile in accessing the APs is when a specific vehicle is exactly equidistant from both APs. The ad hoc communication delay, \( d_{AH} \) (s), which is the time it takes the vehicle to forward the data to any of the APs, depends on the average vehicle population density, \( \rho \) (car/km), the average vehicle velocity, \( v_h \) (m/s) and the average delay per hop, \( d_{hop} \), (s) on different parts of the highway. If we assume that the arrival of vehicles on the highway is Poisson, which is a common assumption in the literature [48], then the inter-distances between vehicles on the highway with average density \( \rho \) have exponential distribution with an average \( 1 / \rho \). Therefore, \( d_{AH} \) can be written as

\[
d_{AH} = (1 - e^{-\rho r})(A/2) * d_{hop} + e^{-\rho r}(A/2) v_h,
\]

(3.21)
where \((1-e^{-\rho r})\) is the probability that the packet is forwarded via wireless communications while \(e^{-\rho r}\) is the probability that the vehicle carries the packet itself, because it cannot find any next-hop vehicle within its range. If the packet is being forwarded via V2V communications, then \((A/2 / r)\) is the number of hops that are needed to cover a distance equal to \(A/2\), which constitutes the forwarding delay when multiplied by \(d_{hop}\). Note that the rationale to approximate the division of the distance by transmission range \(r\) to obtain the number of hops is the packet forwarding mechanism that we consider in which the forwarding vehicle delegates the packet forwarding responsibility to the farthest vehicle in its look-up table. On the other hand, \((A/2 / v_h)\) is the delay the packet experiences when it is carried along a distance equal to \(A/2\). We assume that vehicles are equipped with digital maps that contain the locations and street names as well as the average density and average velocity of vehicles in each street, which are the main factors for decision making in our proposed VHN with V2V communications. A good commercialized example of such digital maps can be found in [46].

One of the parameters, which needs to be determined in (3.21), is \(d_{hop}\). In [66] the performance of the IEEE 802.11 DCF when the traffic is uniformly distributed is analytically characterized for the non-saturated case and \(d_{hop}\) is estimated with high accuracy between

\[
\begin{align*}
\text{Fig. 3.4: Two consecutive APs with distance } d &= A + W \\
\end{align*}
\]
two very close upper bounds and a lower bound. We give more details on the computation of $d_{hop}$ in Section 3.6. In the next subsections, the VHO decision-making problem when V2V communication is included as an alternative is studied with both cost minimization and transmission time minimization goals.

### 3.5.2 Calculation of cost and transmission time

Assuming the computed ad hoc communication delay satisfies the application’s delay requirement, the communication cost when using only WLANs and ad hoc communications, interchangeably called WLAN plus ad hoc in the rest of the chapter, ($c_3$) can be written as

$$c_3 = T_u r_w c_w,$$  \hspace{1cm} (3.22)

where $T_u$ is the usage time which is the total amount of time the connection to APs remains established. This usage time is given by,

$$T_u = (N_w - 1) \left( \frac{W + A}{v} - d_{Ah} \right) + \frac{W}{v} + \left( \frac{x}{v} - d_{Ah} \right),$$  \hspace{1cm} (3.23)

where $N_w$ is the maximum integer value such that

$$\left[ (N_w - 1) \left( \frac{W + A}{v} - d_{Ah} \right) + \frac{W}{v} \right] r_w \leq b_f,$$  \hspace{1cm} (3.24)

and $x$ is obtained from

$$b_f - \left[ (N_w - 1) \left( \frac{W + A}{v} - d_{Ah} \right) + \frac{W}{v} \right] r_w = \left( \frac{x}{v} - d_{Ah} \right) r_w.$$  \hspace{1cm} (3.25)

Note that due to definition of $d_{Ah}$ as the worst case delay, (3.23) approaches the lower bound of usage time, $T_u$. Alternatively, by setting $d_{Ah}=0$, only in the second term at the right hand side of (3.23), we can compute the upper bound usage time too. Also, note that the transmission time ($T_{w+Ah}$) when WLAN plus ad hoc is being used is different from $T_u$ and equals...
So far we have discussed the cases where only cellular networks, WLAN plus cellular or WLAN plus ad hoc are used. Now, we turn to the case where the combination of all three access technologies is selected as the best way of data transmission. This type of transmission called *WLAN plus cellular plus ad hoc* is particularly beneficial when $d_{AH}$ does not satisfy delay requirements but the vehicle is willing to have the most economical set of access networks.

### 3.5.3 WLAN plus cellular plus ad hoc

The idea is that for the given maximum tolerable delay, the distance $A'$ which corresponds to an ad hoc communication delay, equal to the maximum tolerable delay, is obtained according to (3.21). Then, ad hoc communications is used for distances $A'/2$ in the beginning and $A'/2$ in the end and cellular communications are used for transmission over the rest of A. Note that in this case the number of VHOs is smaller or equal to the case of WLAN plus cellular. Therefore, if $c_2 < c_1$, the cost incurred by using WLAN plus cellular plus ad hoc ($c_4$) will also be $c_4 < c_2 < c_1$. Similarly, if $T_w < T_c$, the transmission time in the case of WLANs plus cellular plus ad hoc will be $T_{c+w+AH} < T_w < T_c$.

When the inter-distances of APs are not fixed and they show up independently, the decision-making vehicle does not count on upcoming APs. Based on its distance to the previous AP it uses ad hoc communications for distances smaller than $A'/2$ and cellular communications for any longer distance. The discussion in the previous paragraph on relative costs and relative transmission times of WLAN plus cellular plus ad hoc compared to WLAN plus ad hoc or only cellular communications still holds. Note that the case in which the delay experienced by a packet is approaching the maximum tolerable delay is one of the scenarios.
that trigger the recomputation of decision-making parameters which includes filtering unacceptable alternatives.

3.6 Performance Evaluation

In this section we compare the performance of the proposed distributed system selection scheme with the case where global knowledge of the network is available. For this purpose, we simulated the network using MATLAB. IEEE 802.11p DSRC WLANs are not expected to be widely deployed in the near future. However, IEEE 802.11p MAC layer is derived from IEEE 802.11 MAC and its physical layer parameters are based on IEEE 802.11a standard with minor changes. IEEE 802.11a supports data rates ranging from 6 Mbps to 54 Mbps depending on the distance between the transmitter and the receiver [67]. We have selected the values of WLAN parameters in our evaluation based on [68]. As a result of these parameter choices, the coverage area of APs is set to 100 meters. Further, the signalling overhead required to perform a VHO is determined based on the assumptions made in [69].

Our choice of cellular system is CDMA20001x-EV [70] with an average data rate of 600 Kbps. The type of applications which are of concern in our study is non real-time non-safety application, e.g., file transfer, non real-time multimedia services, data delivery, etc. In these applications a stream of $b_t$ data bits is required to be transmitted. So, the preferences of users could be transmission cost minimization or total transmission time minimization. Both of these approaches will be investigated in our evaluations. A complete list of the parameters used in the evaluation, including those related to the mobility model and the wireless communications system, are listed in Table 3.2.
Table 3.2: Mobility-related and wireless communication-related parameters used in the simulation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W$</td>
<td>100 m</td>
</tr>
<tr>
<td>$A$</td>
<td>200 m</td>
</tr>
<tr>
<td>Decision-making vehicle’s velocity</td>
<td>5–40 m/sec</td>
</tr>
<tr>
<td>Average velocity</td>
<td>10–25 m/sec</td>
</tr>
<tr>
<td>$\rho$ (Average vehicle density)</td>
<td>2–10 veh/km</td>
</tr>
<tr>
<td>$r_w$</td>
<td>6 Mbps</td>
</tr>
<tr>
<td>$r_c$</td>
<td>0.6 Mbps</td>
</tr>
<tr>
<td>$b_{VHO}$</td>
<td>8.8 Mb</td>
</tr>
<tr>
<td>$c_w$</td>
<td>1 Unit</td>
</tr>
<tr>
<td>$c_c$</td>
<td>4 Unit</td>
</tr>
<tr>
<td>$r$ (Transmission range)</td>
<td>100 m</td>
</tr>
<tr>
<td>MAC layer</td>
<td>IEEE 802.11 DCF</td>
</tr>
<tr>
<td>Max. contention window</td>
<td>32</td>
</tr>
<tr>
<td>Propagation model</td>
<td>Two Ray Ground</td>
</tr>
<tr>
<td>Simulation time for ad hoc delays</td>
<td>10000 seconds</td>
</tr>
<tr>
<td>Beacons frequency</td>
<td>2 beacons/sec</td>
</tr>
<tr>
<td>Beacon size</td>
<td>512 bits</td>
</tr>
</tbody>
</table>

We first investigate the performance of the VHO decision-making algorithm, for the scenarios with fixed and random AP inter-distances. We analyze both the transmission times as well as transmission costs versus the velocity of the decision-making vehicle. There exist three networking possibilities, namely accessing and remaining only in the cellular network, using VHO between WLAN and cellular giving priority to WLAN over cellular when available (we call this WLAN plus cellular), and using WLAN connection as well as V2V communication (we refer to this as WLAN plus ad hoc). Note that in the more realistic case of random inter-distances, when WLAN plus cellular is being used, our proposed algorithm makes decisions based on prior knowledge of average distance between consecutive APs. Hence, the decisions may be different from those made based on global knowledge of the network.

Figs. 3.5 and 3.6 demonstrate the VHO policies that should be used to achieve time minimization and cost minimization, respectively. We have presented the results for average
AP inter-distances of 300 meters for all networking possibilities, and for the case in which global knowledge of the network is available. The graph for global knowledge stands for the case where decisions are made with the aid of the network in a centralized manner. For obtaining each of the values in the graph we used Monte-Carlo method. Based on this method, the placement of APs has been randomly generated using a Poisson distribution which requires that the AP inter-distances follow a negative exponential distribution with an average equal to 300 meters, and the transmission time or cost is computed with respect to this placement deterministically in a centralized manner. Each of the presented values in the graph is averaged over 30 individual computations. The analytic and global knowledge results for WLAN plus cellular scenario show a high degree of similarity, thereby verifying our approach.

Fig. 3.5: Comparison of transmission time performance in the proposed algorithms, when $b_t = 1$ Gb
Fig. 3.6: The cost of communication using different VHO policies for the scenario similar to Fig. 3.5

The figures demonstrate the system selection policies that should be used to achieve time minimization and cost minimization. As observed in the figures, the choice of VHO to a WLAN when available (i.e., WLAN plus cellular) seems generally a viable strategy in low speed vehicular communications. However, as the speed of a communicating vehicle increases, the rate of VHOs increases, and the vehicle spends less time in each WLAN coverage area transmitting less traffic to that AP. In the case of WLAN plus cellular access scheme, the extra cost of performing VHO makes staying on the cellular network a more efficient choice. This no VHO strategy applies only to a limited scenario of very high speed vehicles (around 35 m/s).

Fig. 3.7: The required number of VHO for the proposed vehicular communication strategies
Fig. 3.8: Upper and lower delay bounds for $d_{hop}$ [66]

In Fig. 3.7 the number of VHOs needed for sending $b_t = 1$ Gb is shown versus the velocity of the vehicle. This result, further confirms our intuition of optimal VHO decision-making in the previous paragraphs. As it can be observed from Fig. 3.7, the number of VHOs will increase almost exponentially as the speed increases, thereby making VHO a less attractive choice in high speed vehicular networks.

The simulation results in Figs. 3.5-3.7 provide comparisons of accessing only cellular system, WLAN plus cellular network, or WLAN plus ad hoc. The data transmission time or communication cost performance of the latter case, depicted for different average vehicle densities, outperforms other networking possibilities. Although the case of WLAN plus ad hoc shows a relatively better performance, the introduced ad hoc delay may not be tolerable for many applications. Therefore, we will now introduce further simulation scenarios to investigate the performance of ad hoc communication mode in detail. More specifically, since the delay problem becomes more challenging when multi-hop communications are used over larger distances, we study the dependence of ad hoc delays ($d_{ahc}$) incurred by multi-
hop communications on the AP inter-distances. As stated before, to determine parameter $d_{\text{hop}}$ in (3.21), we used the results of [66] in which the performance of the IEEE 802.11 DCF when traffic is uniformly distributed among all nodes is analyzed for the non-saturated case. According to queuing theory [65], in a non-saturated network the average arrival rate of the input traffic is smaller than the average rate at which the traffic is being served, which is equivalent to successful transmission to the next-hop node.

In this study $d_{\text{hop}}$ is characterized with a high accuracy between two very close upper bounds and a lower bound, as depicted in Fig. 3.8. The lower bound and the upper bounds are obtained by using queuing theory [65], [71] and considering that the arrival of packets is unknown and follows a general distribution, i.e., a $G/G/1$ system [66], as follows.

$$E[T_s] \leq d_{\text{hop}} \leq 2E[T_s],$$

(3.27)

$$d_{\text{hop}} = E[T_s] + E[R] \leq E[T_s] + \frac{E[T_s^2]}{2E[T_s]} = T_{UR},$$

(3.28)

where $T_s$ is the service time and $d_{\text{hop}}$ is comprised of both the waiting time in the queue and the service time. Parameter $p$ in Fig. 3.8 is the probability that one node encounters collisions when it transmits, which is equal to

$$p = 1 - (1 - p_t)^{n-1},$$

(3.29)

where $n$ is the number of the nodes that are contending for the wireless media and can potentially cause collision for each other, and $p_t$ is the transmission probability of each node in any time slot which depends highly on the packet traffic of the network.

As it is observed in Fig. 3.8, the lower bound and the upper bounds are very close, and as a result, $d_{\text{hop}}$ can be estimated with high accuracy. Since the collision probability, $p$, is small for non-saturated case, among two higher bounds, $T_{UR}$, which is tighter for smaller $p$’s
is taken into account. According to Fig. 3.8, if $p$ is kept smaller than or equal to 0.1, $d_{hop}$ will remain below 30 ms. Based on this observation and the assumptions in [66], for any given average vehicle density on the highway, the average value of $n$, the number of contending vehicles, can be determined. By taking the calculated $n$ into account and using (3.29), the maximum background packet traffic in the network that yields $p = 0.1$ is obtained which is necessary to make sure $d_{hop}$ remains below 30 ms. The background packet traffic in the network is constantly kept below the calculated values. For different AP inter-distances, $A$, ranging from 0 meter to 600 meters with the vehicle located at equal distances from the APs, the average $d_{AH}$ versus $A$ is obtained via both simulation and analysis by employing (3.21), which is shown in Fig. 3.9. As it is observed in this figure, the simulation results agree well with the results of analyses. The parameters utilized in the simulations are given in Table 3.2. Based on the graphs in Fig. 3.9, for any required delay and given average density, $A'$, which is the distance over which ad hoc networking is allowed, is obtained. Clearly, for distances greater than $A'$ the use of cellular communications is inevitable.

![Fig. 3.9: Ad hoc delay as a function of inter-distance of APs](image)

Analysis, $\rho_{av} = 2$ veh/km  
Simulation, $\rho_{av} = 2$ veh/km  
Analysis, $\rho_{av} = 4$ veh/km  
Simulation, $\rho_{av} = 4$ veh/km  
Analysis, $\rho_{av} = 6$ veh/km  
Simulation, $\rho_{av} = 6$ veh/km  
Analysis, $\rho_{av} = 8$ veh/km  
Simulation, $\rho_{av} = 8$ veh/km  
Analysis, $\rho_{av} = 10$ veh/km  
Simulation, $\rho_{av} = 10$ veh/km
Finally, in order to study the case of WLAN plus cellular plus ad hoc, the maximum tolerable delay, which specifies the distance over which ad hoc data forwarding is allowed, should be determined. Different end-user multimedia categories for a variety of multimedia services were investigated in [60] and it turned out that the maximum allowable one-way transmission delays for a relatively large number of multimedia services is either 1 or 5 seconds. Thus, we set the allowable maximum tolerable delay, i.e., the allowable two-way transmission delay in our simulations to 2 or 10 seconds, respectively. For these maximum tolerable delays and for different average vehicle densities, which is the other parameter affecting ad hoc communications, we compared the transmission times and costs with only cellular and WLAN plus cellular cases. Other parameters remain the same as before. The results for different velocities of the vehicle are shown in Fig. 3.10 and 3.11. It is observed that as long as the speed is not too high, the case of WLAN plus cellular plus ad hoc communications yields remarkable performance improvements over the other cases. As clear in Fig. 3.9, for average vehicle densities equal to or larger than 10 veh/km and the average AP inter-distances of 300 meters, the maximum ad hoc delay stays lower than the pre-determined maximum tolerable delays which yields the same results as the case of WLAN plus ad hoc. That is why in our simulations we have obtained the results for more challenging situations in which average densities are smaller than or equal to 10 veh/km.
inefficiency stems from the fact that when traveling at high speeds it is very likely that a speed, such legacy access technology selection schemes will not be optimized anymore. The cost of communication using different VHO policies for the scenario similar to Fig. 3.10

Fig. 3.10: Comparison of transmission time performance in the proposed algorithms

Fig. 3.11: The cost of communication using different VHO policies for the scenario similar to Fig. 3.10

3.7 Summary

A crucial aspect of vehicular networking in a heterogeneous wireless environment is the optimal choice of access technology among available connectivity alternatives for vehicular end-users. Even though many papers have already investigated this problem for mobile users in general, in vehicular networks where nodes are highly dynamic and move at very high speeds, such legacy access technology selection schemes will not be optimized anymore. The inefficiency stems from the fact that when traveling at high speeds it is very likely that a VHO from a larger access network to a newly emerged smaller access network is
immediately followed by another VHO back to the original access network as soon as the vehicular end-user leaves the coverage area of the smaller network. Hence, traveling at high speeds may cause too many VHOs that in turn incur additional costs and larger transmission times due to the VHO signaling overhead. As a result, the optimal system selection or VHO decision-making strategies should be based on several factors including the available capacity of each access technology, the cost of transmitting traffic in that network and the speed of the vehicle, among others. In the first part of this chapter we have studied how optimal VHO strategies should be adapted to the mobility profiles and service requirements of vehicular end-users in a vehicular heterogeneous network comprised of WLAN and cellular access networks and how such strategies should be designed to be operative on every vehicle in a distributed manner. We have shown that in order to minimize the cost of communications or alternatively to minimize the communication time, performing VHOs is an appropriate choice in lower speeds, whereas it would be better to avoid VHOs and stay in the cellular network at higher speeds. More specifically, the results obtained have provided meaningful decision-making criteria to be employed by individual vehicular end-users that are closely related to the status of the user and the available connectivity options at the decision-making instant.

Another connectivity alternative is the possibility of multi-hop ad hoc communications amongst vehicles that can be considered as an independent system selection option. In the second part of the chapter, we have demonstrated that if V2V communications are also allowed, exploiting the full degrees of freedom, i.e., the combination of WLAN plus cellular plus ad hoc networking, outperforms any other networking strategies in terms of transmission times and transmission costs. Despite its better performance, the WLAN plus cellular plus ad
hoc networking scheme can potentially suffer from larger delays caused by the multi-hop ad hoc communications. As a result, for delay-sensitive applications the delays of this scheme should be calculated and compared with the maximum allowable delays of the applications as a part of the decision-making process to make sure the delay requirements of the users are not violated.

In order to give an overview of the proposed VHO strategies and highlight the main differences between the proposed VHO strategies and existing VHO techniques, the main challenges that were addressed by the proposed VHO strategies and the corresponding features employed in the proposed VHO strategies to address them are listed in Table 3.3. Furthermore, one of the main novelties of the proposed scheme is that it takes advantage of V2V communications or ad hoc networking to remove the need to access costly cellular network at the cost of V2V delays, i.e., the combination of V2V and V2R communications, as opposed to only V2R communications in most existing VHO management studies. Considering such a connectivity alternative necessitates the calculation of distances over which V2V delays remain below delay constraint.

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>VHO strategies should be adapted to mobility profiles of vehicular users, because mobility-unaware VHO strategies (majority of existing works) incur too many VHOs causing additional costs and transfer times</td>
<td>In the proposed VHO strategies, every vehicle makes decisions based on its mobility profile and its preferences in terms of transfer time or cost minimization: we formulated total transfer times and costs, based on which VHO decisions are made, as direct functions of mobility profiles of vehicular users</td>
</tr>
<tr>
<td>VHO decisions should be formulated based on tangible performance metrics, i.e., transfer cost or transfer time</td>
<td></td>
</tr>
<tr>
<td>Challenge</td>
<td>Feature</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Infrastructure-based solutions based on a centralized location server with global knowledge incur deployment and maintenance costs and are not scalable</td>
<td>In the proposed VHO strategies, VHO decisions made by vehicular users in a distributed way, no infrastructure involved</td>
</tr>
<tr>
<td>Existing works are based on periodic decisions at the beginning of discrete time intervals</td>
<td>In the proposed VHO strategies, vehicles continuously seek the optimal access network in an event-based manner</td>
</tr>
</tbody>
</table>
Chapter 4: Hybrid Multi-Technology Routing in Heterogeneous Vehicular Networks

4.1 Introduction

In recent years, various wireless access networks employing different wireless access technologies have been deployed to provide end-users with a wide range of services. As service providers increase the coverage of their access networks, it is more likely that there are overlaps between the coverage areas of different access networks. Such a heterogeneous network provides various connectivity alternatives for end-users. End-users moving at vehicular speeds can benefit from such a rich set of connectivity options to access the Internet for a wide range of Internet protocol (IP)-based applications such as email, content delivery, file download, gaming services, IP telephony, and multimedia streaming. In these applications, vehicular nodes equipped with multi-technology radios need to establish an efficient route to the most appropriate attachment point using the most appropriate set of intermediate hops. Attachment points are the interfaces to the core network, such as BSs in the case of WiMAX or cellular networks, or APs in the case of WLANs, e.g., IEEE 802.11 a/b/g/p WLANs. Numerous single-technology routing protocols, in which packet forwarding over all of the intermediate hops is based on a single wireless technology, have been proposed for vehicular environments. In this chapter, in order to take advantage of the available heterogeneous environment, we study multi-technology routing protocols in which the combinations of different wireless technologies for packet forwarding over intermediate hops are considered.

In a heterogeneous environment it is important to differentiate the problem of designing routing protocols from the problem of optimal access network selection, which has already
been extensively studied in the literature [7-12, 92]. These studies consider the case where end-users are directly covered by several attachment points and decisions should be made to select the most appropriate attachment point for receiving service. However, in a more general case an end-user may not be directly covered by any attachment point or even if an attachment point is available in a single hop, other alternate attachment points could still be preferred. In this case, it is necessary to employ a reliable, robust and efficient routing protocol that finds the most appropriate attachment point in a larger neighbourhood and forwards the packets between the end-user and the attachment point.

Relatively few papers have investigated the issue of multi-technology routing in heterogeneous environments, especially for vehicular networks. In [13] the integration of cellular and WLAN access networks is proposed in which an agent in the cellular network assists the WLAN communications to improve the performance of the network. In [14] cellular and WLAN access networks are combined with the aim of QoS provisioning in a ubiquitous environment. The authors of [15] consider a heterogeneous vehicular networking topology in which every end-user can access both WiMAX and WLAN. The end-users’ WiMAX radios are to be registered in one WiMAX BS. The BS predicts, in a centralized manner, the positions of all vehicles based on which it computes the most appropriate routes between any two end-users. In all these studies it is assumed that the access networks with larger coverage areas, e.g., the cellular or WiMAX network, provide global coverage which allows for end-users to directly connect to it at any location any time. Hence, these networks are used as back-up to provide service any time when networks with smaller coverage areas such as WLANs are unavailable. Clearly, as the size of vehicular networks may become extremely large in practice, considering such back-up network may not be realistic. Hence, in
our heterogeneous topology all access networks regardless of the size of their coverage areas are used as independent connectivity alternatives for multi-hop multi-technology packet forwarding. To the best of our knowledge, none of the previous work has considered multi-hop multi-technology routing for vehicular networks.

In this chapter, we consider a vehicular networking environment in which the movements of vehicles are confined by the structure of roads. Since vehicles may move at very high speeds and in different directions, the topology of the network becomes highly dynamic making the design of routing protocols in vehicular environments very challenging. In this regard, many single-technology topology-based and position-based routing protocols have been proposed [16, 17, 30, 33, 34, 38, 44, 53]. The details of topology-based and position-based routing protocols are elaborated in Sections 1.1 and 2.1. The downside of topology-based routing protocols in vehicular environments is that the links are fairly unstable when packets are forwarded over short-range wireless networks such as WLANs. When the transmission range is short relative to the distances that vehicles travel over a round-trip time between the source and destination, it is very likely that some intermediate vehicles in the end-to-end route get out of each other’s transmission ranges and the route fails even before any data packet is sent on the route. Some efforts have been made to take the stability of routes into account in the process of establishing them [30, 33, 34]. However, when routes are longer than just a few hops, finding stable end-to-end routes becomes very challenging if not impossible, and in sparse situations it is very likely that an end-to-end route may not even exist due to disconnections. So, position-based routing protocols are gaining popularity.
The advantage of position-based routing is that the forwarding of packets does not
depend on the establishment of an end-to-end route. So, this type of routing is a better choice
for highly varying topologies, such as packet forwarding over vehicular networks employing
WLAN technology. The downside of this type of protocols is that the forwarding decisions
are local and without considering real-time network conditions in terms of connectivity and
congestion in other parts of the network. To address these shortcomings, more recent studies
have proposed connectivity-aware routing schemes [38, 44, 53]. However, in these schemes
the connectivity information is pre-determined, and as a result, real-time connectivity and
congestion information regarding the parts of the network that are going to be visited in the
future is not available. On the other hand, in these schemes the general approach for selecting
the most connected route is to make intermediate vehicles report metrics such as average
number of neighbors, minimum number of neighbors and average density of neighbors.
Finally, the route with the maximum value of any of these metrics is considered as the most
connected route. However, these approaches may not be accurate enough, because even
though all these metrics intuitively result in the most connected route, the connectivity in the
context of position-based routing is defined as the probability that no disconnection exists
along the route. A disconnection is the state where no next hop vehicle can be found along
the route, thereby making the communication impossible. To select the route with the
maximum connectivity, an approach for calculating the connectivity according to the
aforementioned definition is required.

The main idea of our work in this chapter is to integrate the advantages of topology-
based and position-based routing into a unified scheme. Based on the fact that the route
instability problem of topology-based routing can be largely overcome using long-range
wireless networks such as WiMAX or cellular networks, we propose the Hybrid Multi-Technology Routing (HMTR) protocol, which takes a hybrid approach for forwarding packets. In MHTR, topology-based routing is used for forwarding packets over more stable links available in long-range networks, and position-based routing scheme is used for forwarding packets over highly variable links in short-range networks. To determine the stability of a link, we propose a link stability logic which is based on the relative mobility of the vehicles forming the link and the delay requirements of the application involved. As a part of HMTR, route selection logic is suggested to prioritize candidate routes based on QoS metrics, network and user preferences and the connectivity of routes. In this regard, we propose a novel microscopic approach for calculating the connectivity of routes on the basis of the connectivity observations of individual vehicles along the routes. To facilitate service delivery in the studied vehicular heterogeneous environment, we also introduce a novel network architecture to address issues such as Authentication, Authorization and Accounting (AAA) in a multi-operator scenarios. To the best of our knowledge HMTR is the first multi-technology, multi-operator hybrid routing protocol for vehicular communications.

The rest of the chapter is organized as follows. In Section 4.2, the network topology is introduced, which can be comprised of an arbitrary set of wireless access networks. In Section 4.3, the HMTR routing protocol is explained. We elaborate on the mechanisms and logics designed for HMTR including the route selection logic and the link stability logic in Section 4.4. The proposed route connectivity is detailed in Section 4.5. The performance of HMTR with respect to its different routing possibilities is evaluated in Section 4.6. Section 4.7 summarizes the chapter.
4.2 Network Topology

4.2.1 Assumptions

As in most other studies [16, 17, 30, 33, 34, 38, 44, 53] we assume that all vehicles are equipped with GPS receivers which can provide position, velocity and time information. Also, all vehicles can obtain roadmap information via digital maps installed in them. Other than the road topology, digital maps also include the ranges of speed and average vehicle densities in every street or highway in the map [46]. Every vehicle can be equipped with one or more digital radios each using a different wireless access technology. We assume that multiple radios onboard a vehicle can be operated simultaneously with no interference to each other; e.g., they employ different frequency bands. Furthermore, we assume that every vehicle has an updated list of all of its one-hop neighbors. For instance, in the case of WLAN access networks this is accomplished by having all WLAN radios periodically broadcast beacon messages in their one-hop neighbouring areas reporting their positions. It is further possible to estimate the velocity vector of other WLAN-enabled vehicles by analyzing their consecutive beacon messages. Every WiMAX radio is also able to obtain an updated list of all other WiMAX-enabled vehicles in its range [72-75], e.g., via the BS.

4.2.2 The topology

We keep the network topology general by assuming that the network topology could be comprised of various access networks. Even though both tight and loose coupling approaches in terms of the architectural design for integrating various access networks are possible [12, 76], in our topology, we select the loose coupling approach for two main reasons:

1. Any of the attachment points, i.e., BSs or APs, may be owned by a different service provider which has its own AAA policies.
2. Since vehicular networks are usually very large networks composed of a number of smaller access networks, their scalability is of great concern. To make the network scalable, we are interested in a topology that requires as few changes as possible in the architecture of readily available access networks in the deployment phase. Since most of access networks have been designed to have Internet access via gateways included in their core networks, loose coupling calls for the minimum required changes in integrating the access networks.

To give an example, the proposed topology when comprised of WLAN, WiMAX and cellular access networks is depicted in Fig. 4.1, in which the larger ellipses, hexagons and smaller ellipses represent the coverage areas of the WiMAX BSs, cellular BSs and WLAN APs in access networks 1, 2 and 3, respectively. The topology we introduce here is different from most commonly used topologies in the literature from two viewpoints:
1. In most previous studies, the access networks with larger coverage areas and usually costlier service such as WiMAX and cellular are used as back-up connectivity alternatives which take over the packet forwarding responsibility when smaller coverage networks fail. This assumption often time requires that a tight coupling approach is used in which the network with larger coverage makes system switching decisions. On the contrary, in our topology any of the available wireless technologies is considered as an independent connectivity alternative which is in accordance with the loose coupling approach.

Fig. 4.1: Network topology
2. Ad hoc networking in a heterogeneous setting can be advantageous when vehicles are not covered by any attachment points or in the case where desirable access networks are available but are out of range. Even though only a few papers in the literature have studied the possibility of ad hoc networking in a heterogeneous environment [8], these papers employ ad hoc networking only as a means for forwarding data to the attachment points that are pre-selected. In our topology we consider ad hoc communications as an independent connectivity alternative which enables us to take the appropriateness of both the possible multi-hop routes and the attachment points into account as opposed to only the attachment points.

The optimum route might consist of links of subscribers of different operators. Each operator or service provider has its own AAA server which interacts with the gateways and AAA servers of other access networks to verify identity, accept or reject access and for billing purposes. As depicted in Fig. 4.1, some of the attachment points in the local core networks of different service providers have dual functionalities of acting as access nodes as well as Internet gateways for connecting the local core network to the Internet. In our topology, both WLAN and WiMAX communications provide ad hoc packet forwarding capability, while cellular communications only provide direct connections from vehicles to cellular BSs and therefore can be only used as the last hop. The possibility of using ad hoc communications over WiMAX radios is explained in more details in Subsection 4.5.1.
4.3 Hybrid Multi-Technology Routing

The mechanism of HMTR can be divided into three different phases including *disseminating a request packet, route selection and returning a reply packet.*

4.3.1 Dissemination of a request packet

Any end-user wishing to establish a connection with an attachment point generates a *request* packet and broadcasts it in the network using all of its available radios, e.g., simultaneous over its WLAN and WiMAX radios if it is so equipped. Any intermediate vehicle that receives the request packet rebroadcasts it on all of its available radios no matter which radio the packet was received on until an attachment point receives the request packet. Since the potential recipients of request packets could be any of the available attachment points, the use of an *anycasting* mechanism is inevitable. In anycasting the same IP address is shared among all attachment points in the network for addressing request packets. This IP address translates into the same ID for all the attachment points to which the request packets are destined. In this routing protocol, to mitigate packet flooding effects we employ several methods to limit the propagation of request packets in the network. One way is to restrict the propagation of request packets to a limited geographical area. Other methods are detailed in Section 4.4. Note that for wireless technologies which do not support ad hoc networking, e.g., cellular network, the request packets are directly forwarded by the onboard radio to the corresponding attachment point. The radio cannot be used at that point if the vehicle is not within the coverage area of any attachment point.

As mentioned in Section 4.1, in HMTR we use a hybrid packet forwarding approach in which topology-based routing scheme is used for forwarding packets over stable links, and position-based routing is used for forwarding packets over unstable links. A link is
considered stable if it is expected to stay valid before the expiry time of the request packet, which is determined by the application requesting the route. The logics employed by intermediate vehicles in HMTR to evaluate the stability of links for a received packet is explained in subsection 4.4.2. To implement the hybrid packet forwarding approach in HMTR, the intermediate radios that use position-based routing include their *locations* in the header of the request packet, whereas the radios that use topology-based routing include their *IDs* in the request header, before rebroadcasting the packet.

### 4.3.2 Route selection

If the request packet is received by more than one attachment point and (or) the same attachment point receives the request packet from different intermediate nodes, more than one route exists and the most appropriate one must be selected. For this purpose, two approaches are possible: *centralized* and *distributed*. In the centralized approach a route selection centre is included in the topology to which all the attachment points forward their received request packets. The centre then selects the most appropriate route according to a route selection logic and generates a *reply* packet containing the selected route to be sent back to the requesting vehicle. In the distributed approach, every attachment point generates a *reply* packet and sends it back and it is up to the requesting vehicle to select the most appropriate route based on the route selection logic. Note that every attachment point also has a unique IP address known to the core network. This IP address is included in the reply packet to start a unicast connection between the attachment point and the requesting vehicle.

Some disadvantages of the centralized approach are as follows:
1. The centralized approach requires the deployment of a new network element, whereas we are interested in solutions that minimize the required changes in the structure of existing networks and impose no additional deployment cost.

2. In the centralized approach the route selection decisions are made only based on one-way traversal of packets in the network, i.e., from vehicles to attachment points. However, reply packets may experience different QoS in the case of asymmetric routes or when reply packets go through different intermediate nodes due to topology changes. Therefore, more accurate route selection decisions can be made based on reply packets that reflect the network conditions on both going and returning ways.

As a result, in our protocol we take the distributed approach. The route selection logic that we incorporate in HMTR is explained in subsection 4.4.1.

4.3.3 Returning a reply packet

The reply packet any attachment point generates includes the route its corresponding request packet has come from in the header. In the state-of-the-art position-based routing protocols for vehicular networking scenarios, the route is defined as a sequence of junctions or physical locations \([38, 44, 53]\). Hence, in order to let both position-based and topology-based routing work properly, the route in our protocol is defined as a sequence of junctions and IDs. The IDs are the IDs of the intermediate vehicles that use topology-based routing for forwarding the request packet which are recorded in the header of the request packet. The junctions are the physical road locations across which the request packet was forwarded over the radios that use position-based routing, which is calculated by using the locations of intermediate vehicles that employ position-based routing and the digital map of the road which is available to every node. On the way back to the requesting vehicle, the reply packet
is forwarded toward the next junction in the route using position-based routing in the parts of the route described by junctions. In the parts described by IDs, the packet is forwarded using topology-based routing. By taking junctions into account instead of the locations of forwarding vehicles when using position-based routing, we make the protocol robust to frequent topology changes as the locations of junctions are fixed. **Example:** A typical description of a route is depicted in Fig. 4.2. Vehicle $S$ is the requesting vehicle of the route and the dotted and dashed curves in the figure represent position-based and topology-based parts of the route, respectively. Also, assume that the location and the ID of a given vehicle $I$ are denoted by $L_I$ and $ID_I$, respectively. Then, the header of the request packet that the BS receives includes the sequence of locations and IDs ($ID_S, L_A, L_B, ID_C, L_D, L_E, ID_F$). After receiving the request packet, the BS determines the route to vehicle $S$ as ($ID_F, J_5, J_4, ID_C, J_3, J_1, ID_S$). ■

In parts of the route where the reply packet is forwarded using position-based routing, a **greedy geographic forwarding** mechanism is used to forward the packet towards the next junction in the route. In this mechanism every intermediate vehicle forwards the packet to the neighbor geographically closest to the next intended junction in the route. Note that in our protocol this mechanism is only used for packet forwarding towards the next junction as opposed to the final destination. Due to topology or connectivity reasons, to successfully deliver the packet to destination, in many urban scenarios, the packet may need to be temporarily forwarded farther from the destination [38, 44]. If the packet forwarding vehicle does not find any next hop vehicles to forward the packet due to temporary disconnections, it starts carrying the packet towards the next junction until another vehicle comes into its range. The possibility of packet carrying makes the protocol robust to disconnections in sparse
situations as the packets will not be immediately dumped when a disconnection is detected along the route. As the packet is being forwarded toward the next junction using this mechanism, every forwarding vehicle also checks for the next ID in the route using its other radios and forwards the packet to the vehicle with that ID if it detects it. After the reply packets are received and the most appropriate route is selected by the requesting vehicle based on the route selection logic, the data and acknowledgement packets are sent along the route, respectively.

![Fig. 4.2: A typical description of a route](image)

### 4.4 Hybrid Multi-Technology Routing

#### 4.4.1 Route selection logic

The route selection logic is implemented in two steps. The first step is to rule out the candidate routes that do not meet the QoS or budget requirements of vehicles. In the second step, among the remaining candidate routes, attachment points and vehicles select the most
appropriate route from the operator and subscriber perspectives, which are based on their priorities in terms of utilization and price, respectively. These steps are detailed in the following two subsections.

4.4.1.1 QoS and (or) budget filtering

Any application that requests access to the Internet may have constraints on some QoS metrics such as delay or bandwidth, which are indicated in the request headers. If an application has a delay constraint on the round trip times of its packets, upon the generation of a request packet the requesting vehicle includes the generation time of the packet in a field in the header. Any intermediate vehicle that receives this packet subtracts the generation time from the present time to obtain the travel time. At any point the travel time exceeds the delay constraint of the packet, the packet will be dropped.

Similarly, an application may have a bandwidth constraint. In this case, every intermediate vehicle replaces its available bandwidth in the corresponding field in the header if it is smaller than the current value of the field. This value reflects the available bandwidth in the route that the packet has experienced up to that point. The packet will be dropped if the available bandwidth is smaller than its required bandwidth. Obtaining the available bandwidth, which is also termed achievable throughput or residual capacity (bps) in the literature has already been discussed in many papers [77-79]. In most of these studies, the total channel usage is measured and subtracted from the channel capacity to obtain the free residual capacity. In addition to restricting the propagation of request packets to limited geographical areas as mentioned in subsection 4.3.1, these filtering mechanisms also limit the propagation of request and reply packets in the network.
Other than delay and bandwidth constraints, the requesting vehicle may also have a budget constraint indicated in the request header. After calculating the price of the end-to-end route, the attachment point compares it with the budget constraint and dumps the request packet if the maximum budget is exceeded. Otherwise, the attachment point attaches the price to the reply packet that it sends back to the requesting vehicle. This mechanism also limits the propagation of unnecessary packets in the network.

The total price is the summation of the service price and the packet forwarding prices over registered forwarding vehicles. In all access networks, in order to access an attachment point, the corresponding radios have to be registered with the attachment point. However, to relay packets to other vehicles in an ad hoc manner, some wireless technologies, e.g., WiMAX, may require the intermediate nodes to be registered in the corresponding access networks, whereas other wireless technologies, e.g., WLAN, may not require such registrations. In other words, WiMAX radios may be charged for packet forwarding while WLAN radios may operate in the ad hoc mode for free. In the former group of wireless technologies, the packet forwarding prices should be taken into account in the calculation of the overall price. For this purpose, we suggest the following charging strategy for packet forwarding.

When an attachment point receives a request packet, it acquires the packet forwarding prices of the vehicles in the route which are registered with the attachment points of other service providers. For this purpose, the attachment point queries their corresponding AAA servers for the prices of packet forwarding by the registered vehicles. As a result, if the requesting vehicle selects to use the route comprising those vehicles for packet forwarding, the AAA servers charge the attachment point instead of the registered packet forwarding
vehicles. The attachment point in turn charges the requesting vehicle. Note that the communications to and from the AAA servers take place on the core network via the Internet. Other than the cost of packet forwarding over the registered vehicles, the attachment point also charges the requesting vehicle for the service it requests, i.e., the service cost.

4.4.1.2 Candidate route selection

Operator Perspective: When the same request packet is received and retained by an attachment point from different routes, all of the candidate routes meet the QoS and (or) the budget requirements. Now, the attachment point should select the most appropriate one for which to generate a reply packet. In order to maximize their revenue, service providers need to make sure the network capacity is used at its fullest which is equivalent to maximizing the utilization. To use the capacity of the network efficiently, the situations in which parts of the network become congested while other parts are not being used at all should be avoided by balancing the packet traffic in the network. For this purpose, we propose that attachment points obtain the difference between available bandwidth on each route and the required bandwidth and select the route with the maximum difference value. This way, the selected route will be left with the maximum available bandwidth which in turn maximizes the traffic balancing in the network, thereby minimizing the probability of congestion. We define $U = \{\text{route 1, route 2, \ldots, route n}\}$ as the set of all candidate routes at the attachment point for a given request. If we denote the available bandwidth along route $j$ and the bandwidth required by the application by $BW_j$ and $BW_{req}$, respectively, the attachment point selects the route with the maximum difference value, namely route $k$, as follows

$$\text{route } k = \arg \max_{\text{route } j \in U} (BW_j - BW_{req}). \tag{4.1}$$
Subscriber Perspective: On the other hand, if the requesting vehicle receives more than one reply packets each generated by a different attachment point, it is generally interested in selecting the cheapest route that meets its QoS requirements. Since all the routes selected by attachment points meet its QoS needs, the requesting vehicle simply selects the cheapest option. We define \( U' = \{\text{route 1}, \text{route 2}, \ldots, \text{route } n'\} \) as the set of all candidate routes at the requesting vehicle. If we denote the price of route \( j \) by \( P_j \), the requesting vehicle selects the route with the minimum price, namely route \( k' \), as follows

\[
\text{route } k' = \arg \min_{\text{route } j \in U'} (P_j). \tag{4.2}
\]

Up to this point, several user and network-favored parameters such as QoS requirements in terms of delay and bandwidth or budget on the user’s side and real-time network conditions in terms of congestion on the network’s side have been taken into account in the proposed route selection logic. However, the real-time connectivity of routes, which is pertinent to position-based routing, has not yet been considered. The connectivity of a route is a critical metric particularly when the network is sparse. Because when packets are routed towards disconnected streets, which are very likely in sparse situations, packet forwarding is no longer possible and the packets should be carried which causes much longer delays, thereby increasing the chance of delay requirement violation and packet dropping. In order to take the real-time connectivity of routes into account, we modify the route selection logic as follows. Note that the real-time connectivity is based on more recent vehicular traffic information which is obtained on-the-fly as packets are disseminated in the network, rather than the pre-stored traffic information in the digital maps of vehicles, which may be obsolete and consequently different from present values.
We consider a field in the header of the packets for the connectivity of the route that the packet has come from. How the connectivity of each route is calculated is explained in Section 4.5. For now, we only assume that the connectivity of the route that each packet has come from is known and is stored in the respective field in the packet.

**Operator Perspective:** If we denote the connectivity along route $j$ by $C_j$, in the modified route selection logic the attachment point selects route $k$ as follows

\[
route k = \begin{cases} 
\arg \max_{route \in V} (BW_j - BW_{req}) & \exists route j: 1/\rho_{j_{min}} < R \\
\arg \max_{route \in U} (C_j) & Otherwise 
\end{cases},
\]

where $R$ is the transmission range, $\rho_{j_{min}}$ is the minimum density of vehicles along route $j$ and $V = \{ route j | route j \in U, 1/\rho_{j_{min}} < R \}$. Since the movements of vehicles are confined to streets and the widths of streets are usually much smaller than the radio transmission range of a vehicle, the movements of vehicles can be considered as one-dimensional movements. Therefore, the reciprocal of the minimum vehicle density along a route represents the maximum average distance between the vehicles on that route. Another field in the header of packets has to be considered for the minimum vehicle density along the route, denoted by $\rho_{j_{min}}$ for route $j$. Any intermediate vehicle calculates the vehicle density in its neighboring area and rewrites it in the respective field if its value is smaller than the current value of the field. Based on the periodic beacon messages that a vehicle receives, it knows the number of vehicles in its transmission range. So, by dividing the number of vehicles by the length of its coverage area the vehicle density in its immediate neighbourhood is obtained.

The condition in (4.3) differentiates the situations where the network is sufficiently dense such that at least one connected route can be found from the situations where the network is so sparse that no such route can be found and therefore packets need to be partly
carried by vehicles before they are forwarded. If routes with sufficient levels of connectivity are found, they can be ranked by the operator or user based on the logic in (4.1) or (4.2), respectively. Otherwise, the route with the maximum connectivity is selected, as given by (4.3). Note that disconnections can only occur in the process of position-based routing as the stability of links have already been verified for the parts of the route involved in topology-based routing. Hence, we are only interested in the density of the vehicles participating in the position-based routing.

**Subscriber Perspective:** Similarly, the requesting vehicle selects route $k'$ according to the following modified route selection logic

$$
route \; k' = \begin{cases} 
\arg \max_{route \; j \in V'} (P_j) & \exists route \; j : 1/\rho_{j_{\text{min}}} < R \\
\arg \max_{route \; j \in U'} (C_j) & \text{Otherwise}
\end{cases},
$$

(4.4)

where $V' = \{route \; j | route \; j \in U', 1/\rho_{j_{\text{min}}} < R\}$. As the requesting vehicle is constantly moving and new attachment points become available, it is very likely that after a while the selected route is no longer the most appropriate route. Hence, the fields in the header of packets are updated every time packets are forwarded between attachment points and requesting vehicles to determine if the current route is about to become invalid and a new route needs to be established.

**4.4.2 Link stability logic**

In order to evaluate the stability of a link for a received packet, the period the link is expected to be valid for, i.e., the Link Lifetime (LLT) is calculated and compared to the expiry time of the packet in its header determined by the application. The link is considered stable for the given packet if its LLT is larger than the expiry time of the packet. The air interface of some wireless technologies supports Non-Line-Of-Sight (NLOS) operations. For
instance, the air interface of WiMAX technology has adopted scalable Orthogonal Frequency-Division Multiple Access (OFDMA) technology which supports variable bandwidth sizes between 1.25 and 20 MHz for NLOS operations [80, 81]. If the link between the communicating vehicles is a NLOS link, a two-dimensional circular radio coverage can be considered. In this case, the lower bound of the LLT is taken into account which can be easily calculated by considering the sequence of streets along which the two vehicles leave the circular ranges of each other faster.

In the following, we give a method on how the LLT of a link can be calculated when the wireless technology used over the link only supports Line-Of-Sight (LOS) operations. For any vehicle moving along a street we define leaving borders. A leaving border for a vehicle is a border beyond which the vehicle is considered to be in a new street. As an example, the leaving borders for vehicles $A$ and $B$ are shown in Fig. 4.3. If the two vehicles are in the same street, the time $t$ that takes them to leave each others’ transmission ranges $R$ can be obtained from

$$
\begin{cases}
|V_A t - V_B t + P_A - P_B| = R \; ; moving \; in \; the \; same \; directions \\
V_A t + V_B t + P_A - P_B = R \; ; moving \; in \; the \; opposite \; directions
\end{cases}
$$

(4.5)

where $V_A$ and $V_B$ are the velocities of the vehicles which only take positive values and $P_A$ and $P_B$ are their one-dimensional positions along the street with respect to the moving direction of the vehicle which is calculating the LLT.

It may be the case that one of the vehicles passes its leaving border before the two vehicles leave each others’ transmission ranges. In this case, there is a high chance of link breakage at the corners of the junctions due to the objects blocking the line of the sight, unless the new street has the same direction as the previous one. Hence, we need to obtain the turning probabilities from the mobility model and calculate the average LLTs which may
not be quite accurate. As an alternative, we take the lower bound of LLTs into account. For the two vehicles in Fig. 4.3 we have

\[ \text{The lower bound of LLT} = \min(t, t_A, t_B), \]

where \( t \) is obtained from (4.5), and \( t_A \) and \( t_B \) are the earliest times the current and the previous vehicles get at their leaving borders and are obtained from

\[ t_A = (d_A + r)/V_A, \]

\[ t_B = (d_B + r)/V_B. \]

\( d_A \) and \( d_B \) are the distances between the positions of the current and the previous vehicles to the center of the junctions toward which they are moving, and \( r \) is the radius of the junction.

![Fig. 4.3: Leaving borders and other parameters for the vehicles in subsection 4.4.2](image)

### 4.5 Route Connectivity

As mentioned earlier, the *connectivity* in the context of position-based routing is defined as the *probability* that no disconnection exists along the route. A number of previous papers have proposed methods to calculate the connectivity in different street segments in the roadmap, i.e., the probability that the distances between any two adjacent vehicles in a street segment are smaller than the transmission ranges of vehicles [82-84]. However, all these
studies use a macroscopic approach for calculating the connectivity. In other words, they are all based on average values of vehicle densities and vehicle speeds in different streets of the roadmap which are stored a-priori in the digital maps of vehicles. However, due to the highly variable network topology, these average values are very likely to be different from instantaneous real-time connectivity observations of individual vehicles along the route in terms of the density of neighbors that arrive in and leave their coverage areas. To the best of our knowledge, our work is the first work that proposes a microscopic approach for calculating the connectivity of routes on the basis of the connectivity observations of individual vehicles along the routes.

We define the vehicle connectivity for vehicle \( i \), denoted by \( VC_i \), as the probability that there exists at least one vehicle ahead of vehicle \( i \) and at least one vehicle behind vehicle \( i \) in its transmission range. All vehicles continuously calculate their vehicle connectivities. Also, we have vehicles include their most updated vehicle connectivities in the beacon messages that they periodically broadcast. For any given route, the connectivity of the route is the product of the vehicle connectivities of all the intermediate vehicles along the route using position-based routing. Hence, the connectivity of route \( j \) can be written as

\[
C_j = \prod_{i=1}^{N} VC_i ,
\]

(4.9)

where \( N \) is the total number of intermediate vehicles along route \( j \) using position-based routing. In HMTR, we get every intermediate vehicle that uses position-based routing to multiply the value in the connectivity field of any received packet by its own vehicle connectivity and rewrite the result in the connectivity field of the packet before rebroadcasting it. In the following we explain how \( VC_i \) can be calculated for any vehicle \( i \).
A common assumption in vehicular traffic engineering theory is to consider a normal distribution for the speeds of vehicles in every street [84, 85]. For each street the minimum and maximum allowable speeds to be included in the normal distribution of that street are available in the digital map. In this chapter, we take the same approach in that we assume that when a vehicle arrives at a street, it takes a fixed speed which remains the same during its residing time in that street. The fixed speed is randomly selected according to the normal distribution of the street. On the other hand, it is widely accepted that in free-flow conditions, in which streets are not congested and vehicles can move as fast as they want, any fixed point on the roadside observes Poisson arrivals of vehicles [48, 83, 84]. Hence, since vehicles are supposed to move at fixed speeds, they also observe Poisson arrivals of other vehicles in their transmission ranges. A typical scenario of vehicles moving on both sides of a street is depicted in Fig. 4.4.

![Fig. 4.4: Different flows of vehicle arriving in the range of Vehicle i](image)

In Fig. 4.4, vehicle $i$ is moving at speed $v_i$. The arrivals of three independent flows of vehicles are distinguishable by vehicle $i$. The first flow corresponds to the vehicles that are arriving in the transmission range of vehicle $i$ in the opposite direction and from the front. We denote the arrival rate of this flow of vehicles by $\lambda_o$. The second flow corresponds to the vehicles that are moving in the same direction as vehicle $i$ and their average speeds are
greater than $v_i$. Therefore, they arrive in the range of vehicle $i$ from behind. The third flow corresponds to the vehicles moving in the same direction as vehicle $i$ with average speeds smaller than $v_i$ which arrive in the range of vehicle $i$ from the front. We denote the arrival rates of the vehicles moving in the same direction with greater and smaller speeds than $v_i$ by $\lambda_{sg}$ and $\lambda_{ss}$, respectively.

According to our definition the $VC_i$ is the probability that there exists at least one vehicle in transmission range $R$ ahead of it and at least one vehicle in transmission range $R$ behind it. In this chapter in order to calculate this probability, we use queuing theory [65, 71] to model distance $R$ ahead of vehicle $i$ and distance $R$ behind vehicle $i$ with two $M/D/\infty$ queues. The justification of Poisson arrivals of vehicles in the transmission range which is equivalent to the arrivals of customers in the queues was already discussed. The reasoning behind considering a deterministic distribution for the service time is based on our previous assumption regarding the fixed speeds for vehicles. Since every arriving vehicle in the transmission range of vehicle $i$ has a fixed speed $v$, its residing time in the range, which is equivalent to the service time of customers in the queues, equals $R / (v_i + v)$ if it has arrived in the opposite direction or equals $R / |v_i - v|$ if it has arrived in the same direction. Note that we assumed that the speed always takes positive values. Having observed this, in our modeling we use the simplifying assumption that the speeds of the flows of vehicles arriving in the opposite direction, in the same direction with greater speeds and in the same direction with smaller speeds are fixed and equal to their average speeds denoted by $v_o$, $v_{sg}$ and $v_{ss}$, respectively. Note that every vehicle can calculate both the average arrival rates and average speeds of different flows of vehicles in its range based on its observations. Also, the reason we considered an infinite number of servers for the queues is the fact that every vehicle starts
receiving service immediately upon its arrival in the transmission range. Note that the arrivals in the transmission range are mapped onto the arrivals in the queue. The queuing system model is depicted in Fig. 4.5.

![Diagram showing the queuing system model with \( \lambda_i \) as the arrival rate, \( t_{\text{service}} = R/(v_i + v_o) \) for service times, and \( P_{\text{in}}(n) \) and \( P_{\text{out}}(n) \) for the probabilities of customers residing in the back and front queues, respectively.

Fig. 4.5: Distances \( R \) ahead and \( R \) behind vehicle \( i \) modeled as 2 \( M/D/\infty \) queues

In the suggested queuing system model, even if one of the queues does not exist, the arrivals of customers in the other queue and their service times are not affected which shows the independence of the queues. As a result of their independence, the \( VC_i \) equals the probability that at least one customer resides in the queue in the back multiplied by the probability that at least one customer resides in the queue on the front. Hence, if we denote the probability that \( n \) customers reside in the queue in the back by \( P_b(n) \) and the probability that \( n \) customers reside in the queue on the front by \( P_f(n) \), respectively, \( VC_i \) can be written as

\[
VC_i = \sum_{n=1}^{\infty} P_b(n) \sum_{n=1}^{\infty} P_f(n).
\]  

(4.10)

Since \( \sum_{n=0}^{\infty} P_b(n) = 1 \) and \( \sum_{n=0}^{\infty} P_f(n) = 1 \), (4.10) can be written as

\[
VC_i = (1 - P_b(0))(1 - P_f(0)).
\]  

(4.11)
We denote the probabilities that $n$ customers belonging to different flows reside in the queues in the back and on the front by $P_{bo}(n)$, $P_{bsg}(n)$, $P_{bss}(n)$ and $P_{fo}(n)$, $P_{fsg}(n)$, $P_{fss}(n)$, respectively. For instance, $P_{bsg}(n)$ corresponds to the customers in the queue in the back arriving in the same direction with greater speeds. Considering that the arriving flows are independent, (4.11) can be written as

$$VC_i = (1 - P_{bo}(0)P_{bsg}(0)P_{bss}(0))(1 - P_{fo}(0)P_{fsg}(0)P_{fss}(0)).$$

(4.12)

In an $M/D/\infty$ queue with a fixed service time $t_s$, the probability $P_n(t)$ that there exist exactly $n$ customers in the queuing system at time $t$ is equal to the probability that exactly $n$ customers arrived between time $t - t_s$ and time $t$. Since the arriving customers have a Poisson distribution with arrival rate $\lambda$, the probability that $n$ customers arrived in a time interval of length $t_s$ is obtained from

$$P_n(t) = \frac{(\lambda t_s)^n}{n!} e^{-\lambda t_s}$$

(4.13)

which is independent of $t$ and holds for any $t > t_s$. Thus, for any $M/D/\infty$ queue the probability that $n$ customers reside in the queue, $P(n)$, equals

$$P(n) = \frac{(\lambda t_s)^n}{n!} e^{-\lambda t_s}.

(4.14)

By setting the arrival rate and the service time to the corresponding values for any of the flows in (4.12), vehicle connectivity of vehicle $i$ can be obtained as follows

$$VC_i = (1 - e^{-\lambda_t s} e^{-\lambda_{bsg} t_s} e^{-\lambda_{bss} t_s})^2 = (1 - e^{-(\lambda_t s + \lambda_{bsg} t_s + \lambda_{bss} t_s)})^2.$$

(4.15)

Thus, $VC_i$ for vehicle $i$ and $C_j$ for route $j$ can be obtained.
4.6 Performance Evaluation

4.6.1 Preliminaries

To evaluate the performance of HMTR, we consider a simplified network topology comprised of only WLAN and WiMAX access networks. The use of WiMAX digital radios is becoming more popular due to their high data rates which support broadband communications and their long transmission ranges which provide a better coverage compared to WLAN radios. The initial WiMAX standard published in 2004, IEEE 802.16 standard [72], was aimed for fixed end-users. Later on, IEEE 802.16e standard [73] was published in 2006 which provided mobility support to end-users moving at speeds of up to 120 km/hr. IEEE 802.16e provides data rates up to 15 Mbps and transmission ranges up to 10 km.

An important limitation of IEEE 802.16e standard is that it only supports direct communications from BSs to end-users, which reduces coverage areas due to transmission power constraints and path loss. In order to extend the coverage area outside the ranges of BSs, a new draft standard, IEEE 802.16j was approved by the IEEE-SA Standards Board in 2006 [74] which is based on IEEE 802.16e standard and extends the coverage by using multi-hop relaying. In the initial drafts of IEEE 802.16j, the relaying nodes are fixed nodes acting as small scale BSs, requiring that they are enabled with some of the functionalities of BSs. However, in more recent versions these limitations are addressed and multi-hop communications can be carried out over mobile nodes. IEEE 802.16j standard was approved by IEEE-SA Standards Board in 2009 as an amendment to IEEE 802.16 standard [75].

Despite recent developments of WiMAX networks, IEEE 802.11-based WLAN networks will still be used in local environments and continue their growth to become more
ubiquitous and to challenge WiMAX networks in larger areas. The FCC allocated 75 MHz spectrum to DSRC [1] at 5.9 GHz frequency band to V2V and V2R communications in 1999. DSRC was standardized in the draft IEEE 802.11p [2], a variant of IEEE 802.11a standard adjusted for low overhead operations, which supports data rates up to 6 Mbps and transmission ranges up to 300 m. In vehicular settings it is mostly assumed that WLAN radios comply with DSRC 802.11p standard.

Note that the reason we did not include cellular networks in our evaluation model stems from the fact that cellular technology does not support multi-hop packet routing. In other words, since both WLAN and WiMAX technologies provide ad hoc packet forwarding capabilities, they are better options for evaluating the performance of our proposed routing protocol compared to cellular technology which only provides direct connections from end-users to cellular BSs and therefore can only be used as the last hop of the route. It is worth mentioning that the possibility of using cellular communications as the last hop can be easily included in our evaluation model without adding much complexity.

4.6.2 Simulation settings

We evaluate the performance of HMTR via simulation. The road topology we use in the simulation is a grid layout derived from a real street map in the TIGER database [56] from US Census Bureau. The street layout is depicted in Fig. 4.2. In our scenario we simulated the mobility of vehicles and the wireless communications between them separately. For simulating the mobility of vehicles we used SUMO street traffic simulation package [57]. To simulate wireless communications in our scenario, we developed an event-based network simulator using C++. To the best of our knowledge, none of the commonly used network simulators have implemented WiMAX communications with multi-hop relaying on mobile
nodes yet. The outputs of SUMO which include the positions of all vehicles at every time step of the simulation runtime are then used by our network simulator as inputs.

In SUMO, every street is assigned minimum and maximum speeds according to the digital map, and a functionality defines whether the street is a *plain* street, a *source* street or a *sink* street. For any given number of vehicles, every vehicle is randomly injected into one of the source streets. When it reaches a sink street, it is removed from the network and randomly regenerated in another source street, and this procedure continues over the simulation runtime. In order to generate realistic vehicular mobility traces, SUMO supports right-of-way rules at junctions, traffic regulations and traffic lights. Also, additional weights are assigned to different streets to make them more or less attractive for vehicles when they arrive at junctions. We assigned the weights proportional to the average vehicle densities in the road topology stored in the digital map. We observed that after 2000 seconds the errors between the vehicle densities obtained in SUMO and the vehicle densities stored in the digital map became smaller than 5% of the stored values for all the streets in the simulation area. Thus, we start sending packets in the network after 2000 seconds. All vehicle traces during the simulation time are saved in a log file, which is then used in the network simulator.

We have selected the values of WLAN parameters in our evaluation based on [66, 68]. In [66], the performance of the IEEE 802.11 DCF under uniformly distributed traffic among all nodes is analyzed for the non-saturated case. In our simulation scenarios, we define the packet traffic generation in a way that the network is constantly non-saturated. The parameters used in the mobility model and the WLAN parameters are depicted in Table 4.1 and Table 4.2, respectively. The values of the parameters used in the simulation of WiMAX
radios are selected similar to the simulation scenarios in [86] which in turn are based on the discussions in WiMAX forum [87] and readily available deployments of WiMAX device manufacturers, e.g., [88] (Table 4.3). Since the focus of our study is the evaluation of the proposed routing protocol, we disable the WiMAX Adaptive Modulation and Coding (AMC) feature in our simulation scenario. We consider a fixed data rate and also a fixed transmission range adapted from the simulation scenario investigated in [86].

Table 4.1: Mobility-related parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average length of streets</td>
<td>500 m</td>
</tr>
<tr>
<td>Number of vehicles</td>
<td>200 ~ 600</td>
</tr>
<tr>
<td>Average velocity</td>
<td>15 ~ 105 km/h</td>
</tr>
<tr>
<td>Simulation runtime</td>
<td>20,000 s</td>
</tr>
</tbody>
</table>

Table 4.2: WLAN-related parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmission range</td>
<td>250 m</td>
</tr>
<tr>
<td>Radio model</td>
<td>Two ray ground</td>
</tr>
<tr>
<td>Traffic model</td>
<td>CBR over 25 random</td>
</tr>
<tr>
<td>CBR rate</td>
<td>63 packets/s</td>
</tr>
<tr>
<td>Data packet size</td>
<td>1 KB</td>
</tr>
<tr>
<td>Beacon size</td>
<td>512 bit</td>
</tr>
<tr>
<td>Beaconsing frequency</td>
<td>2 beacons/s</td>
</tr>
<tr>
<td>Data rate</td>
<td>2 Mbps</td>
</tr>
<tr>
<td>MAC layer</td>
<td>IEEE 802.11 DCF</td>
</tr>
<tr>
<td>Backoff slot time</td>
<td>20 µs</td>
</tr>
<tr>
<td>SIFS</td>
<td>10 µs</td>
</tr>
<tr>
<td>DIFS</td>
<td>50 µs</td>
</tr>
</tbody>
</table>

Table 4.3: WiMAX-related parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel bandwidth</td>
<td>7 MHz</td>
</tr>
<tr>
<td>OFDM symbol duration</td>
<td>34 µs</td>
</tr>
<tr>
<td>Cyclic prefix duration</td>
<td>2 µs</td>
</tr>
<tr>
<td>Frame duration</td>
<td>20 ms</td>
</tr>
<tr>
<td>Data rate</td>
<td>10 Mbps</td>
</tr>
<tr>
<td>Transmission range</td>
<td>2 km</td>
</tr>
</tbody>
</table>

In our simulation scenario we consider Constant Bit Rate (CBR) traffic with Best Effort (BE) service type. Different end-user multimedia categories for a variety of multimedia services were investigated in [60] and it can be observed that the maximum allowable one-
way transmission delays for a relatively large number of multimedia services is either 10 seconds or 1 minute. Thus, we set the allowable message travel times in our simulations accordingly.

In the simulations we assume that all vehicles are equipped with both WLAN and WiMAX radios. Note that due to the higher bandwidths of WiMAX hops, attachment points, according to (4.1), tend to select the routes with the maximum number of WiMAX hops for any given budget. Having observed this, instead of constraining the budget we can alternatively limit the number of WiMAX hops in the routes and obtain the costs for a given maximum allowable number of WiMAX hops in the routes. In our simulation we consider four different routing possibilities: HMTR with WLAN Only where no WiMAX transmission is allowed, HMTR with 1hopWiMAX in which only one WiMAX transmission is allowed in a route, HMTR with 2hopWiMAX in which routes are allowed to have up to two WiMAX hops, and HMTR with WiMAX Only. Note that the reason for considering only these four routing possibilities is that the area in our simulation scenario can be covered by at most three consecutive WiMAX hops. Also, based on the given density of vehicles and the transmission ranges of WiMAX radios in our scenarios, every WiMAX radio can almost always find another WiMAX radio in its transmission range.

4.6.3 Simulation results

The performance metrics that we consider in our evaluations are packet delivery ratio, packet delivery delay and cost. The packet delivery ratio is the ratio of the packets the network layer delivers to its higher layer to the packets generated in the higher layer and passed onto the network layer for delivery. The packet delivery ratio accounts for packet droppings that take place when the travel times of packets exceeds their delay constraints.
We assume that the transmission costs on all WiMAX radios are fixed and equal to 1 unit per every 1 Kb, the transmissions on WLAN radios are free of charge and the service costs are the same for all service providers for the same service. Therefore, in the calculation of total costs we only consider the transmission costs regarding the packet forwarding over WiMAX radios. In practice, service providers price their packet forwarding and services based on business considerations, among which use of bandwidth is but one of the factors. However, we tackle the problem at hand from a system design point of view to minimize the cost for vehicular end-users.

We run three rounds of simulations. In the first two rounds, we study how the use of different number of attachment points in the network affects the performance. In the third round, we consider a more realistic scenario in which only a fraction of vehicles are equipped with WiMAX radios.

In the first round of simulations only one attachment point exists in the network and is placed at the top rightmost junction of the network. The packet delivery ratios and the packet delivery delays for maximum allowable one-way delays of 1 minute and 10 seconds are depicted in Figs. 4.6 and 4.7, respectively. Note that regarding the large number of samples obtained over the simulation time, for the confidence interval of 95 percent around the mean values, the margins of error for the data points displayed in all of the graphs are of the order of $10^{-4}$. 
As expected, the more WiMAX forwarding is involved, the better the routing performance in terms of both delivery ratio and delay. In the routing possibilities that fully or partially rely on WLAN forwarding, in lower vehicle densities, vehicles mostly resort to packet carrying as opposed to packet forwarding as it is less likely for them to find next-hop WLAN-enabled vehicles in their ranges. As a result, in lower densities only those WLAN radios which are in the proximity of the attachment point can succeed to deliver the packets to the attachment point before the delay constraints are exceeded. Note that in the calculation of packet delivery delays, only the delays of the packets that have been successfully delivered are taken into account. As the number of vehicles increases, the chance of finding next-hop WLAN-enabled vehicles increases and WLAN radios mostly rely on packet
For packet forwarding rather than packet carrying. As a result, almost all packets are delivered before the delay constraints are exceeded. This best explains the peaks in the delay graphs. The transmission costs for different routing possibilities are also shown in Table 4.4. Note that as explained before, transmission costs only depend on the packet forwarding over WiMAX radios. Since for the given transmission ranges and the range of the number of vehicles, vehicles can always find a WiMAX-relaying vehicle in their WiMAX transmission ranges, the average transmission costs stay the same for different number of vehicles.

Table 4.4: Transmission costs for 1 attachment point

<table>
<thead>
<tr>
<th>Routing</th>
<th>Transmission Cost (Units/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMTR with 1hopWiMAX</td>
<td>12,600</td>
</tr>
<tr>
<td>HMTR with 2hopWiMAX</td>
<td>21,656</td>
</tr>
<tr>
<td>HMTR with WiMAX only</td>
<td>24,216</td>
</tr>
</tbody>
</table>

Also, for the same reason, we observe that WiMAX radios can always find other WiMAX radios in their neighborhoods to which they have stable enough links for topology-based routing. Hence, when there is no limit on the use of WiMAX hops in the routes as in the **HMTR with WiMAX Only** case, HMTR stands for a pure topology-based routing and the results are comparable to the performance of most state of the art topology-based routing protocols [30, 33, 34]. On the other hand, in the **HMTR with WLAN Only** case it turns out that almost all of the WLAN hops are unstable as a result of their fast movements with respect to their shorter transmission ranges. Therefore, this case stands for pure position-based routing and the results are representative of the state of the art position-based routing protocols [16, 17, 38, 44, 53]. As a result, while it is obvious that **HMTR with WiMAX Only** shows the best performance, the unique feature of our hybrid protocol is the cost tradeoff with respect to the user budget. In other words, the simulation results show that HMTR provides the opportunity to select an intermediate level of performance in terms of delivery ratio and delivery delay.
for a given budget and average vehicle density, whereas in pure position-based or pure
topology-based routing schemes sacrificing the performance or budget may be inevitable in
many scenarios.

In the second round of simulations, we study the same performance metrics when two
attachment points are deployed. We place the attachment points at the farthest possible
distances from each other, i.e., one of them is placed at the top rightmost junction and the
other one at the bottom leftmost junction of the network. Even though the population of
attachment points is different in different vehicular networking scenarios, e.g., urban or
suburban areas or highway environments, the reason for using few attachment points in our
experiment setup is that the performance of the routing protocols can be better studied when
most users are not directly covered by attachment points and multi-hop routing is the only
possible way to access the core network. The results of the simulation are shown in Figs. 4.8
and 4.9. The transmission costs are also shown in Table 4.5. As we expected, the use of two
attachment points yields better performance for the same routing possibilities, which is a
result of the smaller average distance between vehicles and attachment points. Note that in
the presence of two attachment points and for the vehicle densities, transmission ranges and
vehicle speeds given in the simulation scenario, all vehicles can access one of the attachment
points by at most two consecutive WiMAX transmissions. As a result, the \textit{HMTR with
2hopWiMAX} case gives the same results as the \textit{HMTR with WiMAX Only} case.
performance can be affected by the WiMAX penetration rates, no limit on the maximum fraction of the vehicles in the network are equipped with WiMAX radios. Three cases are considered, in which only 50%, 20% or 10% of vehicles are equipped with WiMAX radios, whereas all vehicles have WLAN radios. In order to study how the best achievable performance can be affected by the WiMAX penetration rates, no limit on the maximum

| Table 4.5: Transmission costs for 2 attachment points |
|-------------------------------------------|-----------|
| HMTR with 1hopWiMAX                     | 12,600 Units/s |
| HMTR with 2hopWiMAX                     | 18,113 Units/s |

It is expected that in the deployment phase of WiMAX technology over vehicles, the penetration rate, i.e., the percentage of vehicles equipped with WiMAX radios is small. Therefore, in the third round of simulations, we study a more realistic scenario where only a fraction of the vehicles in the network are equipped with WiMAX radios. Three cases are considered, in which only 50%, 20% or 10% of vehicles are equipped with WiMAX radios, whereas all vehicles have WLAN radios. In order to study how the best achievable performance can be affected by the WiMAX penetration rates, no limit on the maximum
allowable number of WiMAX hops in the routes is imposed. Also, in the simulation setup we assume that only one attachment point exists in the network and is placed at the top rightmost junction of the network. The packet delivery ratios and the packet delivery delays for maximum allowable one-way delays of 1 minute and 10 seconds are depicted in Figs. 4.10 and 4.11, respectively. It is observed in the figures that even when a small percentage of vehicles are enabled with WiMAX capability, the performance is considerably improved. However, if we keep increasing the percentage of WiMAX-enabled vehicles, the improvement obtained becomes less noticeable.

![Fig. 4.10](image1.png) ![Fig. 4.11](image2.png)

Fig. 4.10: (a) Packet Delivery Ratio for maximum delay of 1 min. and different percentage of WiMAX-enabled vehicles, and (b) Packet Delivery Delay for maximum delay of 1 min. and different percentage of WiMAX-enabled vehicles

![Fig. 4.11](image3.png)

Fig. 4.11: (a) Packet Delivery Ratio for maximum delay of 10 sec. and different percentage of WiMAX-enabled vehicles, and (b) Packet Delivery Delay for maximum delay of 10 sec. and different percentage of WiMAX-enabled vehicles
4.7 Summary

To make use of heterogeneous networking environments for vehicular end-users to access the Internet, in this chapter we have proposed a routing protocol, HMTR, for multi-hop vehicular networks. HMTR has been designed to take into account different combinations of wireless technologies in intermediate hops and is generally formed of a combination of topology-based and position-based routing schemes for packet forwarding. Due to the highly dynamic topology of vehicular networks, the packet-forwarding scheme in any successful routing protocol should be adaptable to the rate of topology changes in the network. For this purpose, we have proposed a link stability logic for evaluating the stability of links. In HMTR for a given packet, the position-based routing approach has been used for packet forwarding over unstable links, i.e., over highly variable links whose lifetimes are shorter than the packet expiry time. On the other hand, HMTR has employed the topology-based routing approach for packet forwarding over more stable links that are expected to stay valid before the expiry time of the packet.

Not only should a successful routing protocol for vehicular networks have a stability adaptive packet-forwarding scheme, but it should also meet the user requirements in terms of budget or QoS metrics such as delay and bandwidth and the service provider preferences in terms of network utilizations, while it maintains adequate levels of connectivity. To select the optimal route among the candidate routes, a two-step route selection logic has been employed in HMTR that takes both subscriber and operator preferences into account. In the first step, the routes that do not satisfy the QoS requirements or the budgets of route requesting applications have been excluded. In the second step, among the remaining candidates, the most connected one has been selected in sparse vehicular traffic situations. Alternatively, in
dense enough vehicular traffic situations the most appropriate candidate has been selected with the purpose of packet traffic balancing or cost minimization by the network or vehicular end-users, respectively. In order to assess the connectivity of the routes, we have also proposed a novel scheme to calculate the connectivity of a given route defined as the probability that no disconnection exists along the route. The proposed approach is microscopic. In other words, contrary to the readily available macroscopic approaches that are all based on the average values of vehicle densities and vehicle speeds stored a-priori in the digital maps of vehicles, microscopic approaches are based on the instantaneous real-time connectivity observations of individual vehicles along the route. The proposed scheme for calculating connectivity is general and could be potentially used as a valid metric for connectivity measurement in other similar studies.

Eventually, we have evaluated these effectiveness of HMTR via simulation. Simulation results have shown that HMTR enables us to achieve the best possible performance in terms of delivery ratio and delivery delay for a given budget, whereas in pure position-based or pure topology-based routing schemes sacrificing the performance or budget may be inevitable in many scenarios. Furthermore, we have observed that even when only a small fraction of vehicles are equipped with WiMAX radios, a considerable performance improvement is achieved over the case of only WLAN radios.
Chapter 5: A Reliable Robust Fully Ad Hoc Data Dissemination

Mechanism for Vehicular Networks

5.1 Introduction

According to the World Health Organization report on road traffic injury prevention, 1.2 million people die and 50 million are injured in motor vehicle collisions each year worldwide. One of the important purposes of enabling vehicular environments with wireless capabilities is to improve the safety of passengers on roads, so-called safety applications. In safety applications, the goal is to avoid impending collisions by disseminating early notifications of hazardous situations, so-called emergency messages, e.g., when airbag deploys in case of an accident or when ABS brake system is activated in case of slippery road conditions. V2V communications, which have recently gained a great momentum, are anticipated to enable a wide range of vehicular applications, including safety applications. In classical vehicular safety systems, the traffic and emergency information is detected by the sensors mounted along the roadside, collected and interpreted in computing centers, and distributed to the vehicles by means of radio broadcast stations or cellular networks. In addition to their high cost, these centralized infrastructure-based systems cannot provide detailed local information for vehicles, because of their limited data rates. Moreover, since the vehicles are not communicating directly, the communication delays can be intolerably high for many applications. Consequently, infrastructure-less distributed ad hoc vehicular solutions, so-called VANETs, have gained popularity among researchers recently.

An extensive number of applications in VANETs, particularly safety applications, need a robust, reliable, bandwidth-efficient data dissemination mechanism between vehicles. However, data dissemination is a challenging task in VANETs, due to highly variable
network topologies [3, 89], frequent network fragmentation in low traffic densities and the movement of vehicles which is constrained to pre-defined roads with specific speed limits [5]. In addition to safety applications, some other types of applications, i.e., traffic and road information sharing in Intelligent Transport Systems (ITS) and non-safety applications such as sale advertisements or announcements for marketing purposes, could also be based on data dissemination mechanisms. Throughout the chapter, we refer to the area in which the data is required to be disseminated as the zone-of-relevance.

When supporting safety applications over DSRC/802.11p, we have to take into account the strict requirements of such applications in terms of reliability, which is evaluated by using delivery ratio as the performance metric, and delivery latency, especially for emergency messages such as Forward Collision Warning (FCW) or Electronic Emergency Break Light (EEBL) requiring strict delay bounds. In other words, the data dissemination mechanism should deliver the emergency messages to every vehicle in the zone-of-relevance and should make sure the delivery latency is kept below a threshold, i.e., the emergency messages are delivered to the drivers in a timely manner, to guarantee that the drivers are given enough time to react to the emergency situations, thereby avoiding accidents and saving lives. Otherwise, the envisioned data dissemination mechanism would be useless. A typical chain accident example is depicted in Fig. 5.1 to explain how data dissemination mechanisms can prevent chain accidents and what role the reaction times of the drivers play in accident avoidance. In this example cars \( A \), \( B \) and \( C \) are driving at 30 m/s and there is a 30 meters spacing between any two consecutive cars. The cars can decelerate at 5 m/s\(^2\) when they brake. Also, we assume that the reaction time of the drivers is 1.5 seconds. Car \( A \) detects an emergency at time 0 and brakes. In the first scenario, no wireless enabled collision
avoidance is in place and as depicted in Fig. 5.1 (a) all the three cars will end up in collision, because they only make decisions to slow down when they see the brake lights of the cars ahead of them. However, in the second scenario, demonstrated in Fig. 5.1 (b), as soon as car A detects an emergency, it disseminates the emergency messages informing cars B and C. Hence, even though car B still hits car A, car C will have enough time to stop safely and avoid the accident.

Fig. 5.1: A chain accident example (a) no wireless data dissemination, and (b) wireless enabled data dissemination
While most of the previously proposed data dissemination mechanisms operate successfully along straight roads, they encounter some challenges for disseminating data around intersections. One possible failure scenario occurs when the emergency message is being forwarded across an intersection and it leaves the intersection without making sure that the message has been forwarded towards all the outgoing roads at the intersection. A typical failure scenario is depicted in Fig. 5.2 in which an emergency message has been generated by car *source* and is being forwarded in the zone of relevance. The solid arrows show the data dissemination flow generated form car *source* within the zone-of-relevance. When the message is being forwarded across the first intersection, it leaves the intersection without being forwarded towards the direction in which car *A* is approaching the intersection. As a result, in this scenario, when car *A* arrives at the intersection, the emergency message has already left the intersection and car *A* will not receive the message. In this respect, some recent work has differentiated data dissemination along straight roads (*directional* mode) and data dissemination around intersections (*intersection* mode) [24-29]. In this chapter, we firstly review a number of existing fully ad hoc *directional* mode data dissemination mechanisms and after evaluating them by means of simulation, we select one with the best performance in terms of delay, efficiency and reliability as the *directional* mode mechanism in our integrated data dissemination mechanism (Section 5.2). More details on possible failure scenarios of current *intersection* mode data dissemination mechanisms are given in Section 5.3 that motivate the proposition of our novel *intersection* mode data dissemination mechanism in Section 5.4. We integrate this *intersection* mode solution with the *directional* mode mechanism mentioned above to form our united data dissemination mechanism, which is the main contribution of this chapter. Section 5.5 summarizes the chapter.
5.2 Directional mode data dissemination mechanisms

In simple flooding and proactive algorithms [20, 21], the data is broadcast to all the one-hop neighbours, which in turn stores and rebroadcasts the data. Obviously, these algorithms are not scalable as they generate a large amount of message traffic. Due to the shared wireless channel, this high message traffic causes a significant amount of redundancy and collisions among neighbouring vehicles. This problem is commonly referred as broadcast storm problem [23]. In order to reduce redundancy and collisions, the authors of [22] offer a directional broadcast-based algorithm in which the messages are forwarded only if they have arrived from the right direction. However, in the presence of a high node density (we use the terms node and vehicle interchangeable throughout this chapter) and heavy traffic, the
network is still prone to the broadcast storm problem. One of the seminal data dissemination mechanisms, which effectively addresses the broadcast storm problem as well as the hidden node problem, is Urban Multi-hop Broadcast (UMB) [24]. The directional mode data dissemination mechanism used in UMB is as follows.

Unlike flooding-based protocols, in UMB, in order to cope with the broadcast storm problem, each vehicle forwards the packet only to the node in the farthest distance within its radio transmission range. Note that UMB is a beacon-free mechanism, i.e., it does not require the exchange of location information among neighbour nodes. To mitigate the hidden node problem, a handshake protocol similar to the Request-To-Send (RTS)/Clear-To-Send (CTS) protocol in IEEE 802.11, called Request-To-Broadcast (RTB)/Clear-To-Broadcast (CTB) is used. The RTB packet contains the position of the source node as well as the intended broadcast direction. When a node in the direction of the dissemination receives the RTB, it transmits a jamming signal called a black-burst with the length proportional to the distance between the node and the source. This mechanism ensures that the farthest node sends the longest black-burst. At the end of its black-burst, every node listens to the channel. If it finds the channel idle, it infers that it is the farthest node to the source. So, it replies to the RTB by sending back a CTB packet to the source. The actual message and an Acknowledgment (ACK) are sent afterwards.

Another promising data dissemination mechanism is Data Pouring with Intersection Buffering (DP-IB) [25]. In the directional mode data dissemination mechanism of this scheme, contrary to UMB, vehicles use periodic beacon messages to report their locations, directions and velocities to each other. To mitigate the broadcast storm problem, each
message-forwarding vehicle selects the farthest vehicle in the data dissemination direction in its neighbour list, which is obtained through beaconing, as the next forwarder.

Both UMB and DP-IB have reliable and robust directional-mode data dissemination mechanisms that translate into the successful delivery of the emergency message to every vehicle available in a zone-of-relevance along a straight road. As a result, in order to select one of them as the directional mode solution in our integrated data dissemination mechanism, we compare their performance in terms of latency considering that in most data dissemination applications, in addition to reliability, the low-latency delivery of the messages is also of utmost significance. To compare the performance of UMB and DP-IB in terms of latency, we formed a simulation system. The simulation is based on a 4 kilometer two-lane road. Note that latency is defined as the total time needed for delivering the emergency message to all of the vehicles in the 4 kilometer zone-of-relevance. In our simulation scenario, we assume that the vehicle closest to the left hand side edge of the zone-of-relevance has detected an emergency event and is therefore responsible for forwarding the message throughout the zone-of-relevance. The parameters used in the simulation system including those related to the mobility model and the wireless communications system are listed in Table 5.1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highway length</td>
<td>4 kilometer</td>
</tr>
<tr>
<td>Vehicle density</td>
<td>5 ~ 40 vehs/km per lane</td>
</tr>
<tr>
<td>Average speed</td>
<td>15 ~ 35 m/sec</td>
</tr>
<tr>
<td>Simulation time (used for computing delivery ratio)</td>
<td>2000 sec</td>
</tr>
<tr>
<td>Emergency message size</td>
<td>2500 bit</td>
</tr>
<tr>
<td>Transmission range</td>
<td>100 m</td>
</tr>
<tr>
<td>MAC Layer</td>
<td>IEEE 802.11</td>
</tr>
<tr>
<td>Max. Contention Window</td>
<td>32</td>
</tr>
<tr>
<td>Parameter</td>
<td>Value</td>
</tr>
<tr>
<td>----------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Background safety traffic</td>
<td>10 kbps</td>
</tr>
<tr>
<td>Data rate</td>
<td>1 Mbps</td>
</tr>
<tr>
<td>Beacon interval</td>
<td>2 ~ 5 beacons/sec</td>
</tr>
<tr>
<td>Beacon size</td>
<td>512 bit</td>
</tr>
<tr>
<td>Radio Model</td>
<td>Two Ray Ground</td>
</tr>
</tbody>
</table>

For the case in which the node density is 20 vehicles per kilometer per lane, and the vehicles are sending two beacons per second, the latency of the two schemes for different average speed of vehicles is depicted in Fig. 5.3. Each result was obtained by taking the average value from 30 simulation runs. The reason that the latency decreases when the average speed of vehicles increases is that at some points in time, due to fragmentation the vehicles should carry the messages until they find the next node to forward the message to. Clearly, in such cases when the message-carrying vehicles move at higher speeds, the messages are delivered to all the vehicles in the zone-of-relevance in shorter times. As observed in Fig. 5.3, UMB has lower delays than DP-IB. One reason is that since in DP-IB every vehicle forms its neighbour list based on the beacon messages that it periodically receives from other vehicles, the accuracy of the neighbour list depends on the beaconing frequency. In other words, the next forwarder that a vehicle chooses from its neighbouring list might have already left the transmission range of the vehicle. One way to cope with this problem is to increase the beaconing frequency. A higher beaconing frequency corresponds to more frequent updates of the neighbour list. As this problem becomes more critical when the average speed of the vehicles increases, we simulated our system for the average vehicle speed of 35 m/s (Fig. 5.4). Although the latency decreases as the beaconing frequency increases, it is still greater than that of UMB. This is due to the fact that contrary to UMB, DP-IB is prone to the hidden terminal problem since it uses the two-way hand-shake mode of the IEEE 802.11 MAC standard, and also the fact that the beaconing itself imposes some
extra traffic and causes more delays in the network. On the basis of the results demonstrated in this section, we select UMB as the directional mode solution of our integrated data dissemination mechanism. Next, we develop a suitable mechanism for data dissemination around intersections.

Fig. 5.3: The comparison between UMB and DP-IB when employed on a straight highway (for 20 vehs/km per lane and 2 beacons per second in DP-IB)

Fig. 5.4: The latency decreases when the beacon interval in DP-IB increases, but still remains greater than UMB (for avg. speed of 35 m/sec)
5.3 Intersection mode data dissemination mechanisms

DP-IB and UMB rely on repeaters installed at the intersections to forward the messages towards the intersecting roads. Another work that relies on static nodes in intersections to help relay data is [28] in which messages are forwarded to the static nodes when there exists no vehicle in the out-going road of interest. The static nodes store the messages and transmit them when nodes become available. However, we are looking for an infrastructure-independent fully ad hoc solution. In [29] an approach similar to UMB is selected for data dissemination. However, the main difference is that a two-way handshake is used as opposed to a four-way handshake, i.e., use of RTB and CTB, for selecting the next vehicle to retransmit. In this work, the same mechanism used in directional-mode data dissemination is also suggested for intersection-mode data dissemination, which obviously does not consider the LOS obstructions at intersections. Also, it suggests the use of packet carrying to deal with sparse situations that does not address the failure scenarios at intersections when no retransmitter is found. To the best of our knowledge, the only fully ad hoc data dissemination mechanism that explicitly addresses data dissemination at intersections is Ad hoc Multi-hop Broadcast (AMB) which is proposed by the authors of UMB in their other work [26]. This is why we consider the intersection mode mechanism in AMB as a baseline mechanism to compare our proposed intersection mode mechanism with.

AMB has the same directional mode data dissemination as UMB; however, it is an extension of UMB in the sense that instead of repeaters, an ad hoc solution is used at the intersections. In AMB an intersection region is defined starting from $R/2$ meters before and extending to $R/2$ meters after the intersection, where $R$ is the radio transmission range. When a message is being forwarded towards an intersection, the first vehicles that receives the
message in the intersection region, which is referred to as the \textit{HUNTER} vehicles, is responsible to find the vehicle closest to the center of the intersection. The radius of the intersection region is set to $R/2$ meters, because the \textit{HUNTER} vehicle tries to select the vehicle closest to the center of the intersection and its transmission range should cover all the areas closer to the center of the intersection than itself. In order to select the closest vehicle to the center of an intersection, the vehicles set the lengths of their black-bursts proportional to reverse of their distance to the center. Once the vehicle closest to the center of the intersection receives the emergency message, it is then responsible for forwarding the message towards the intersecting roads. Since RTB packets contain the intended broadcast direction, a new RTB packet should be sent for each direction.

Even though we already gave a general example to show how data dissemination mechanism could possibly fail at intersections in Section 5.1, in this section we give more detailed failure scenarios to explain why such a failure happens. More specifically, we elaborate on some scenarios in which the AMB mechanism fails. The first scenario is depicted in Fig. 5.5 (a). Let node $A$ be the \textit{HUNTER}. The moving directions of the vehicles are also specified in the figures. Node $A$ forwards the message to node $B$ which is the closest vehicle in its LOS to the center of the \textit{intersection}. However, node $B$ is already out of the intersection. Note that what we refer as the \textit{intersection} is the area in gray in Fig. 5.5 (a) that is different from the \textit{intersection region} defined earlier in the AMB mechanism. When node $C$ or $D$ arrives at the intersection, node $B$ that was responsible for forwarding the message towards the intersecting roads has already left the intersection region and neither of them will be notified.
Another failure scenario is shown in Fig. 5.5 (b). In this scenario, node $A$ forwards the message to node $B$, which is inside the intersection this time. However, node $B$ will leave the intersection before node $C$ comes inside its transmission range, thereby leaving node $C$ un-notified. Another scenario is shown in Fig. 5.5 (c). In this scenario node $A$ delegates the message forwarding responsibility to node $C$ which is the closest node to the center of the intersection. However, node $C$ is leaving the intersection and since it has no LOS to node $D$, there is no way to notify node $D$. If node $A$ had instead delegated the message forwarding responsibility to node $B$, since node $B$ is about to enter the intersection, it would have notified node $D$. In all of these failure scenarios, the message leaves the intersection region without making sure that it has been disseminated towards all the intersecting road segments. In the next section, we propose an effective solution to this problem.
Fig. 5.5: (a) $A$ forwards the message to $B$ which is out of the intersection, (b) $A$ forwards the message to $B$ which is inside the intersection, but leaves the intersection before $C$ arrives at its transmission range, and (c) $A$ had better forward the packet to entering node $B$ rather than leaving node $C$.

5.4 Proposed enhanced intersection mode data dissemination mechanism

In this section, we present our proposed Enhanced Intersection mode Data Dissemination (EIDD) mechanism, which is fully ad hoc in its operations and highly robust. The idea is to keep the emergency message in the intersection long enough to ensure that the message is forwarded to all the intersecting road segments. The procedure is that when a vehicle in the intersection region receives the message, it attached a number of bits equal to the number of out-going roads, other than the one that the message has arrived from, to the header of the message. For instance, if the crossroad is composed of two perpendicular roads,
three bits are added to the header of the message. Considering that vehicles are all equipped with digital maps that include the road topology, they know the number of out-going roads and they use more bits if the intersection is more complex. Each of the bits corresponds to the dissemination of the message to one of the out-going roads. The bits are initially set to zero and their order is determined according to the direction that the message enters the intersection. Since every vehicle is assumed to be equipped with a GPS, it can simply identify the direction from which it has received the message. For instance, for the case with three added bits if the message is coming to the intersection from the west, the three bits correspond to the out-going roads on the north, east, and south, respectively. The vehicle is then responsible for forwarding the message to all the out-going roads, and at the same time if it finds a vehicle closer than itself to the center of the intersection, it delegates the forwarding responsibility to that vehicle. Whenever the vehicle responsible for forwarding the message receives an acknowledgement from any of the required out-going roads, it sets the corresponding bit to one. Delegating the forwarding responsibility to the node closest to the center of the intersection continues until all the three bits are set to one or the message expires. In order to increase the chance of delivering the packet to all the directions and minimize the delay, we define the intersection region starting from $R$ meters before and extending $R$ meters after the intersection. In Fig. 5.6, it is demonstrated how EIDD overcomes the failure scenario discussed in Fig. 5.2 of Section 5.1.

In Fig. 5.6, as soon as car $B$ receives the message, it adds three zero bits to the header of the message to account for the dissemination of the message to the out-going roads on the north, east, and south. After adding the three bits, car $B$ sends the message to car $C$ that is closer to the center of the intersection delegating the message forwarding responsibility. Car
C can immediately find car D and car E in the north and east out-going roads, respectively, which would in turn disseminate the message further in the zone-of-relevance. Hence, it sets the first two bit to one after receiving the acknowledgements from the two vehicles. However, since it has not yet found any vehicle in the out-going road on the south, it starts carrying the message and at the same time check to see if any vehicle closer to the center of the intersection than itself becomes available. Eventually, car A in the out-going road on the south arrives in the transmission range of car C to which it forwards the message. After all the bits are set to one, the message is no longer needed in car C and can be safely removed from its buffer.

![Diagram](image)

**Fig. 5.6:** How EIDD overcomes the failure scenario in Fig. 5.2

To compare the performance of our proposed EIDD mechanism with the intersection mode data dissemination mechanism used in AMB, we formed another simulation system. The zone-of-relevance is shown in Fig. 5.7, with three secondary roads running vertically intersecting with one primary road running horizontally. The length of the primary road is again 4 kilometers and we assume that the vehicle closest to the left hand-side edge of the zone-of-relevance has detected the emergency event. Also the directional mode data dissemination mechanism employed along the straight roads is that of UMB and the vehicle
density in the primary road is set twice that of the secondary roads. Other parameters are the same as before (Table 5.1).

![Diagram](image)

**Fig. 5.7:** The primary road intersected by three secondary roads

Since EIDD overcomes the failure scenarios discussed before in an effective way, we expect that EIDD performs better than AMB in terms of reliability and robustness, particularly at lower vehicle densities where disconnections are more likely to happen. To study this issue, we obtained the delivery ratio, which is defined as the percentage of the vehicles that finally gets notified in the simulation time, in terms of vehicle density. As it is observed in Fig. 5.8, the delivery ratio of EIDD is much higher than AMB, and is equal to 100% for approximately every given vehicle densities, except when the average distances between vehicles are smaller than the transmission range. As the vehicle density increases, the delivery ratio of AMB reaches a maximum and decreases afterwards due to a larger number of collisions which results in the scenarios where the *HUNTER* vehicle leaves the intersection without having successfully forwarded the message in all the outgoing directions. However, the delivery ratio of EIDD appears to remain 100% even at higher vehicle densities, thereby guaranteeing the scalability of EIDD.
Fig. 5.8: The delivery ratio of EIDD is considerably higher than AMB and approximately 100%.

For the case where vehicle density is 20 vehicles per kilometer per lane, for the simulation runs in which the message is delivered to all the vehicles, we obtained the average latency of AMB and EIDD for different average speeds. It can be observed in Fig. 5.9 that both EIDD and AMB show very similar average latencies in such scenarios.

Fig. 5.9: EIDD and AMB have approximately similar delays.
5.5 Summary

Many applications in vehicular networks need the data to be disseminated from a source vehicle to a large number of vehicles in the network. Although many data dissemination mechanisms have been previously proposed by the research community that operate successfully along straight roads, some challenges and failure scenarios still remain unsolved when the data is being disseminated at an intersection and should be forwarded to all the intersecting road segments. As a result, the data dissemination along straight roads, i.e., directional mode, should be differentiated from the data dissemination at intersections, i.e., intersection mode. In the first part of this chapter, we have introduced a number of directional mode data dissemination mechanisms and compared their performance in terms of delivery latency and reliability that are two of the commonly accepted performance evaluation metrics for data dissemination mechanisms by means of simulation. It has been demonstrated that the directional mode data dissemination mechanism employed in UMB outperforms its peers.

In the second part of the chapter the challenges and failure scenarios that occur during the data dissemination at intersections have been studied that call for an intersection-aware design of the data dissemination mechanisms. Many of the readily available data dissemination mechanisms rely on repeaters installed at the intersections to forward the messages towards the intersecting roads. However, due to the reasons explained in the chapter such as the deployment and maintenance costs of infrastructure, we have only focused on infrastructure-independent fully ad hoc solutions in this chapter. Hence, we have investigated fully ad hoc intersection mode data dissemination mechanisms and concluded that in order to overcome such failure scenarios, a successful directional mode data
dissemination mechanism should be integrated with a successful intersection mode data dissemination mechanism. Consequently, we have proposed a united data dissemination mechanism integrating the directional mode data dissemination mechanism with the best performance obtained in the first part of the chapter, UMB, with our novel reliable robust fully ad hoc enhanced intersection mode data dissemination mechanism, EIDD. Finally, we have evaluated our united solution in terms of robustness, reliability and scalability and compared its performance with other plausible data dissemination mechanisms. The results obtained confirm the effectiveness of EIDD that suggests that EIDD could be a good candidate for data dissemination in VANETs.

Currently we are studying the use of infrastructure on the performance of the data dissemination mechanisms in vehicular networks to see if it makes a considerable improvement over fully ad hoc solutions. Furthermore, considering other more complicated applications that call for the use of routing protocols and might impose the co-existence of various access technologies (in the form of a heterogeneous network) is also the subject of our future research.
Chapter 6: Concluding Remarks and Possible Future Directions

In this chapter, we conclude the dissertation by highlighting the contributions of each chapter and summarizing the results. Also, we present a number of the possible directions for further research.

6.1 Summary of accomplished work

In the present thesis, we have proposed application-specific packet routing mechanisms and protocols for various vehicular applications that carefully take the challenges, considerations and specific requirements of each vehicular application into account. More specifically, we have proposed a number of routing protocols and Vertical Hand-Off (VHO) strategies for provisioning non-safety applications in vehicular homogeneous and heterogeneous networks in Chapters 2, 3 and 4. We have developed a novel data dissemination mechanism for addressing safety applications in Chapter 5.

- In Chapter 2, we have presented a connectivity-aware minimum-delay routing protocol for Vehicular Ad hoc Networks (VANET) that employs a novel route selection logic adaptable to the density of vehicles in the network. In order to deal with network disconnections in sparse situations, the routes with higher vehicle densities are prioritized and when the network is dense, less congested routes with minimum delays among routes with enough connectivity level are favored. A target tracking mechanism is included in the protocol to deal with the movement of target vehicles. Also, a decision scheme is proposed for forwarding packets in sparse junctions which aims to minimize end-to-end delays by taking every promising forwarding opportunity into account. The performance of the proposed routing protocol has been compared with the performance of two plausible geographic
connectivity-aware routing protocols for VANETs, Vehicle Assisted Data Delivery (VADD) and Anchor-based Street and Traffic Aware Routing (A-STAR). The obtained results showed that CMGR outperforms A-STAR and VADD in terms of both packet delivery ratio and ratio of dropped data packets. For example, under the specific conditions considered in the simulations, when the maximum allowable one-way transmission delay is 1 minute and one gateway is deployed in the network, the packet delivery ratio of CMGR is approximately 25% better than VADD and A-STAR for high vehicle densities and goes up to 900% better for low vehicle densities.

- In Chapter 3, we have proposed optimal VHO decision-making strategies for optimal selection of access networks at any point in time. A crucial aspect of vehicular networking in a heterogeneous wireless environment is the optimal choice of access technology. This optimal VHO decision in general depends on several factors such as the available capacity of each access technology, the cost of transmitting traffic in that network and the speed of the vehicle, among others. In Chapter 3 we have considered a vehicular heterogeneous network comprised of a Wireless Local Area Network (WLAN) and cellular systems. We have shown that in order to minimize the cost of communications or alternatively minimize the communication time, use of VHO is an appropriate choice in lower speeds, whereas it would be better to avoid VHO and stay in the cellular network at higher speeds. Furthermore, we have demonstrated that if Vehicle-to-Vehicle (V2V) communication is also possible, the combination of WLAN plus cellular plus ad hoc networking outperforms any other networking strategies that we have considered in terms of transmission times and transmission costs.
• In Chapter 4, we have proposed a hybrid multi-technology routing protocol for Internet access in vehicular networking environments. To make packet forwarding adaptable to the rate of topology changes in the network, in the proposed protocol we have used position-based and topology-based routing approaches for packet forwarding over unstable and stable links, respectively. In this regard, we have proposed a link stability logic for evaluating the stability of links. Among the candidate routes, the most appropriate one has been obtained by using a two-step route selection logic. The first step is to exclude the routes which do not satisfy the Quality of Service (QoS) requirements or the budgets of route requesting applications. In the second step, among the remaining candidates, the most connected one has been selected in sparse vehicular traffic situations. Alternatively, in dense enough vehicular traffic situations the most appropriate candidate has been selected with the purpose of packet traffic balancing or cost minimization by the network or vehicles, respectively. In this regard, a novel scheme has been proposed to calculate the connectivity of a given route. Simulation results have shown that the proposed protocol enables us to achieve the best possible performance in terms of delivery ratio and delivery delay for a given budget, whereas in pure position-based or pure topology-based routing schemes sacrificing the performance or budget may be inevitable in many scenarios.

• In Chapter 5, we have considered the issue of data dissemination in VANETs. First, we have introduced a number of directional mode data dissemination mechanisms and showed by means of simulation that the directional mode data dissemination mechanism used in Urban Multi-hop Broadcast (UMB) outperforms its peers. Then,
we have brought up a number of challenges and failure scenarios that call for our novel fully ad hoc enhanced intersection mode data dissemination mechanism. We have evaluated our united solution in terms of robustness, reliability and scalability. The results obtained have showed that our proposed data dissemination mechanism could effectively address the difficulties encountered in vehicular networking environments and recover from the common failure scenarios in vehicular roadmaps. EIDD has been compared with another plausible fully ad hoc data dissemination mechanism with intersection awareness. The obtained results showed that EIDD outperforms its peer in terms of delivery ratio in different vehicle densities. Under the specific conditions considered in the simulations, the delivery ratio of EIDD is 100% for approximately all the vehicle densities that shows its robustness to dense situations where the network becomes congested and packet collisions are most likely.

6.2 Future directions

In the following, some interesting directions for extending the work presented in the present dissertation are introduced.

1. **Optimum number of Gateways (GW) and repeaters to be added in a homogeneous network and their optimal locations:** the focus of the proposed routing protocol in Chapter 2 has been to achieve the best possible performance in the readily available platforms based on the needs of specific vehicular applications. In other words, we assumed that we have no control over the structure of the underlying network. Also, in Chapter 5 we have assumed that no infrastructure is available. Obviously, even though the proposed solutions are the
best possible solutions that can be obtained for the given platforms, they do not always meet the requirements of all different types of vehicular applications of concern. One possible interesting direction to extend the proposed work is to modify the structure of the network in hand, by adding GWs, and repeaters with the purpose of guaranteeing the support of QoS requirements for any given vehicular application. Two important questions to answer in this regard are the optimum number of required GWs or repeaters to be added and their optimal locations.

2. **Optimum number of GWs and repeaters to be added in a heterogeneous network and their optimal locations:** due to the same reason explained in point 1, the study on the VHO strategies proposed in Chapter 3 and the routing protocol proposed in Chapter 4 can be extended to address more stringent QoS requirements. In heterogeneous networking environments, adding new attachment points, i.e., APs, BSs or any other type of infrastructure, is more challenging compared to homogeneous networks. Because, the main question that we should answer in this context is the most appropriate type of attachment point in terms of access technology to be deployed in a given area. This question should be answered even before the optimal number of attachment points and their optimal locations are calculated. The most appropriate type of attachment point at a given area is expected to be closely related to the relative density of readily deployed attachment points of each access technology and the QoS requirements of the applications of concern, among other variables.
3. **Request-adaptive data dissemination for context-aware applications:** in almost all the data dissemination mechanisms including the one we proposed in Chapter 5, the emergency message should be disseminated in a pre-determined neighborhood. In other words, the objective of a data dissemination mechanism is to deliver the same message to every vehicle in the neighborhood of concern. However, in many newly emerged applications such as context-aware and location-aware applications, the observations of individual vehicles located in a neighborhood of concern are required to be collected. Some examples of the required observations could be a specific number of non-overlapping observations or at least one observation on every street segment in the neighborhood of interest. For this purpose, the message should be updated on the fly by individual vehicles as it is being disseminated in the neighborhood of concern. Another possible direction for future research is to extend our proposed data dissemination mechanism to make it adaptable to the needs of different context-aware applications. In the extended mechanism, the intermediate vehicles should be able to update the messages with their own observations or the results of their local processing.
Bibliography


[65] X. Chao, M. Miyazawa, and M. Pinedo, Queueing Networks. 1999.


