Geo-Routing and Transport Protocols for Efficient and Reliable Vehicular Communications

Konstantinos Katsaros

Submitted for the Degree of Doctor of Philosophy in Electronic Engineering from the University of Surrey

Institute for Communications Systems
Faculty of Engineering and Physical Sciences
University of Surrey
Guildford, Surrey GU2 7XH, U.K.

March 2015

© Konstantinos Katsaros 2015
I would like to dedicate this thesis to my loving parents.
Statement of Originality

This thesis and the work to which it refers are the results of my own efforts. Any ideas, data, images or text resulting from the work of others (whether published or unpublished) are fully identified as such within the work and attributed to their originator in the text, bibliography or in footnotes. This thesis has not been submitted in whole or in part for any other academic degree or professional qualification. I agree that the University has the right to submit my work to the plagiarism detection service TurnitinUK for originality checks. Whether or not drafts have been so-assessed, the University reserves the right to require an electronic version of the final document (as submitted) for assessment as above.

Konstantinos Katsaros
December 2015

Email: K.Katsaros@surrey.ac.uk
www: http://info.ee.surrey.ac.uk/Personal/K.Katsaros/
Summary

Intelligent Transportation Systems (ITS) apply Information and Communication Technologies (ICT) to improve safety and efficiency as well as the passenger experience in modern transport systems. It is envisaged that dynamic vehicular networks, particularly, Vehicular Ad-hoc Networks (VANETs) based on dedicated short-range communications (DSRC) and cellular networks, will be important parts of the future ITS. Unlike traditional communication networks, VANETs are highly dynamic systems resulting in significant reliability issues for the communication protocols. In addition, cellular networks incur notable usage cost. Motivated by this, we investigate efficient and reliable geo-routing and transport protocols aimed at VANETs and VANET/cellular hybrid architectures.

Specifically, first we develop an innovative, unicast, cross-layer, weighted, position-based routing protocol (CLWPR) that takes into account mobility and cross layer information about neighbour nodes. A heuristic algorithm based on analytic hierarchy process (AHP) is employed to combine multiple decision criteria into a unique weight parameter used to select the node to which the packet is forwarded. Comprehensive simulations are performed in realistic representative urban scenarios with synthetic and real traffic. Insights on the effect of different communication and environment parameters are obtained. The results demonstrate that the proposed protocol outperforms existing routing protocols for VANETs, including ETSI’s proposed greedy routing protocol, GyTAR, and AGF in terms of combined packet delivery ratio, end-to-end delay, and overhead.

To efficiently distribute location information, required for the proper functionality of geo-routing, we develop a centralised Location Service. Exploiting the availability of two interfaces (DSRC and LTE) in a hybrid system, we propose separation of signalling and data traffic. The former is transferred over a cellular network and the later over a short range ad-hoc network. For the evaluation of the proposed scheme, we develop an analytical model of the upper bound delay based on stochastic network calculus (SNC) theory. We compare the upper-bounds of three networks, namely a pure short-range ad-hoc network, a pure cellular based on 3GPP LTE and the proposed hybrid with signalling on cellular and data on ad-hoc network. The results of our investigation suggest that hybrid networks can significantly improve performance of vehicular networks in terms of end-to-end delay both for data and signalling traffic.

In the light of these findings, we investigate transport protocols for hybrid networks benefiting from multi-homing support. As Stream Control Transmission Protocol (SCTP) is one IETF standard that supports multi-homing, we develop an analytical model for throughput calculation of a round-trip time (RTT)-aware SCTP variant. Finally, we propose a novel SCTP scheme that takes into account not only path quality but also the cost of using each network. We show that the combination of QoS and cost information increases economic benefits for provider and end-users, while providing increased packet throughput.

Key words: vehicular ad-hoc networks, geographic routing, cross-layer optimization, location service, hybrid networks, transport protocol, end-to-end delay modelling.
**Acknowledgements**

Some will claim that the PhD is a lonely road to walk. I can assure you it was nothing like that. Through this process I had the opportunity to meet a lot of great people, visit beautiful places and learn exciting stuff. Learning (no matter the circumstances or level) opens up the window to a fascinating life.

A lot of people supported me and helped me open this window, to whom I wish to express my gratitude. First of all, I would like to thank my supervisors, Dr Mehrdad Dianati and Prof Rahim Tafazolli, for the help and support that they gave me. Furthermore, I would like to thank current and past member of CCSR for fruitful technical discussions, and the friendships we have developed. In particular I would like to mention Chong, Ralf and Nick, as well as all the administrative staff (Safa, Alison and Chris).

Moreover, as this work was partially funded by the joint EU FP7 research project DRIVE C2X and Huawei, I would like to thank the partners for their support and opportunities they gave me.

Last but not least, I would like to thank my family and particularly Angeliki for her support and understanding all these years.
Contents

List of Figures .......................... xiii
List of Tables ........................... xv
List of Abbreviations .................... xviii

1 Introduction .......................... 1
1.1 Scope ................................ 1
1.2 Challenges and Motivations ........... 3
1.3 Objectives ........................... 9
1.4 Contributions ....................... 10
1.5 Thesis Overview ..................... 12

2 Background and Related Works ........ 15
2.1 Intelligent Transport Systems ........ 15
2.2 Applications for ITS ................ 16
2.3 Related Works ...................... 17
  2.3.1 Routing Protocols ................. 17
    2.3.1.1 Routing protocols taxonomy for VANETs 18
    2.3.1.2 Geographic Forwarding Mechanisms ........ 19
    2.3.1.3 Cross-Layering Designs .................. 27
  2.3.2 Location Services ................. 31
    2.3.2.1 Taxonomy and characteristics ............ 31
    2.3.2.2 Infrastructure-less LS .................. 33
    2.3.2.3 Infrastructure-based LS .................. 34


2.3.3 Transport Protocols ................................................. 34
  2.3.3.1 Single Best Path Selection ................................. 35
  2.3.3.2 CMT Best Path Selection .................................. 36
2.3.4 Propagation Models ................................................. 38
2.3.5 Delay Modelling ..................................................... 40
2.4 Summary ............................................................... 42

3 Cross-Layer Optimised Geo-Routing 47
  3.1 Introduction .......................................................... 47
  3.2 Protocol Design ...................................................... 48
    3.2.1 Mobility related information ................................. 50
      3.2.1.1 Dynamic HELLO Broadcast scheme ..................... 50
    3.2.2 Link quality related information ............................ 52
      3.2.2.1 CSI Function .............................................. 52
    3.2.3 Node utilisation related information ........................ 54
  3.3 CLWPR Operation With an Example ............................... 55
  3.4 CLWPR Weighting Function ....................................... 56
  3.5 Analytic Hierarchy Process (AHP) ............................... 58
    3.5.1 Description of problem as a hierarchy .................... 58
      3.5.1.1 Construction of pair-wise comparison matrix .......... 59
    3.5.2 Calculation of \( f_X \) parameters .......................... 59
  3.6 Performance Evaluation of CLWPR ................................. 60
    3.6.1 Impact of Forwarding Parameters ............................ 61
    3.6.2 Comparison with ETSI-GF, AGF and GyTAR .................. 67
  3.7 Discussion and Summary .......................................... 69

4 Hybrid Network Location Service Architecture 71
  4.1 Introduction ....................................................... 71
  4.2 Evaluated Location Service Network Architectures ............ 72
    4.2.1 Location Service Operation ................................. 73
    4.2.2 IEEE 802.11p-based Location Service Architecture .......... 74
    4.2.3 Hybrid Network Location Service Architecture .............. 75
## 4.2.4 Piggybacking Location Header ........................................... 75

### 4.3 Performance Analysis of Location Service ........................................ 76

#### 4.3.1 Location Service Success Ratio ........................................... 77

#### 4.3.2 End-to-End Delay .................................................. 78

#### 4.3.3 Overhead .......................................................... 79

### 4.4 Summary .............................................................. 79

## 5 Delay Bound Modelling .................................................. 83

### 5.1 Introduction ......................................................... 83

### 5.2 Overview of Stochastic Network Calculus ........................................ 84

#### 5.2.1 Definitions and Notation ............................................... 84

### 5.3 End-to-End Delay Bounds ................................................. 88

#### 5.3.1 Modelling of Arrival Processes ......................................... 89

#### 5.3.2 End-to-End Delay Bounds: Ad-Hoc Network Scenario ......................... 90

#### 5.3.3 End-to-End Delay Bounds: Cellular Network Scenario ....................... 93

#### 5.3.4 End-to-End Delay Bounds: Hybrid Network Scenario ......................... 98

### 5.4 SNC Model Validation and Performance Evaluation ........................... 100

#### 5.4.1 Model Validation .................................................. 101

#### 5.4.2 Performance Evaluation ............................................... 102

### 5.5 Summary .............................................................. 106

## 6 Cost-Efficient Transport Protocol ......................................... 109

### 6.1 Introduction ......................................................... 109

### 6.2 Analytical Model of RTT-aware SCTP ...................................... 110

#### 6.2.1 Original SCTP modelling ............................................ 110

#### 6.2.2 Modified SCTP model .............................................. 112

##### 6.2.2.1 Definition of \( P^{UP} \) ............................................ 112

##### 6.2.2.2 Queueing Delay Calculation ...................................... 113

#### 6.2.3 Evaluation of RTT-aware SCTP model .................................... 114

### 6.3 Cost Model .............................................................. 115

#### 6.3.1 Off-Loading indicator .................................................. 117

### 6.4 Cost-Efficient Transport Protocol ......................................... 118

#### 6.4.1 Architecture for CE-SCTP ............................................. 120

### 6.5 Performance Evaluation of Cost Effective Transport ......................... 121

### 6.6 Summary .............................................................. 123
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 Conclusions</td>
<td>127</td>
</tr>
<tr>
<td>7.1 Conclusions</td>
<td>127</td>
</tr>
<tr>
<td>7.2 Future Work</td>
<td>131</td>
</tr>
<tr>
<td>A Supplementary Results</td>
<td>135</td>
</tr>
<tr>
<td>A.1 Comparison with topology-based routing protocols</td>
<td>135</td>
</tr>
<tr>
<td>A.2 Impact of Carry-n-Forward mechanism</td>
<td>136</td>
</tr>
<tr>
<td>B Publications</td>
<td>139</td>
</tr>
<tr>
<td>Bibliography</td>
<td>141</td>
</tr>
</tbody>
</table>
List of Figures

1.1 How we travel? ................................................. 2
1.2 Why we travel? ............................................... 2
1.3 Impact levels of Intelligent Transport Systems .......... 4
1.4 ITS reference architecture (a), Network and Access Layers (b) .... 8

2.1 Intelligent Transport Systems ................................ 16
2.2 Forwarding Methods without navigation .................. 22
2.3 Forwarding Mechanisms for VANETs ....................... 24
2.4 Local Maximum Example .................................... 26
2.5 Location Service taxonomy .................................. 32
2.6 Propagation in urban scenarios with LOS and n-LOS components .......... 39

2.7 Communication links in urban scenario simulated using NS-3 with and without considering the effect of building .......... 39

3.1 CLWPR forwarding algorithm .................................. 49
3.2 PDR vs. HELLO Interval for low density traffic .............. 51
3.3 Proposed CSI function ........................................ 53
3.4 CLWPR forwarding algorithm example ..................... 54
3.5 AHP hierarchy for CLWPR weighting function ............... 58
3.6 Vehicle traces overlaid on “Unterstrass” area in Zurich .......... 61

3.7 Performance of CLWPR protocol for different $\{c_{L,M}, c_{T,M}, c_{T,L}\}$ parameters ........................................ 64
3.8 Performance of CLWPR protocol for different $\{c_{P,D}, c_{R,D}, c_{R,P}\}$ parameters ........................................ 65
3.9 Performance of CLWPR protocol for different $\{c_{M_a,C}, c_{N,C}, c_{N,M_a}\}$ parameters ........................................ 66
3.10 Performance comparison with V2V connections (Scenario #1) ........ 68
<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.11</td>
<td>Performance comparison with V2I connections (Scenario #1)</td>
<td>68</td>
</tr>
<tr>
<td>3.12</td>
<td>Performance comparison with real scenario with V2V connections (Scenario #2)</td>
<td>69</td>
</tr>
<tr>
<td>4.1</td>
<td>Network model with (a) ad-hoc and (b) cellular Location Service architecture</td>
<td>72</td>
</tr>
<tr>
<td>4.2</td>
<td>Message sequence diagram for Location Service basic functions</td>
<td>74</td>
</tr>
<tr>
<td>4.3</td>
<td>IEEE 802.11p-based location service network architecture</td>
<td>75</td>
</tr>
<tr>
<td>4.4</td>
<td>3GPP LTE-based location service network architecture</td>
<td>76</td>
</tr>
<tr>
<td>4.5</td>
<td>Location Service success ratio under different request rate</td>
<td>78</td>
</tr>
<tr>
<td>4.6</td>
<td>Impact of vehicles’ speed on end-to-end delay and overhead</td>
<td>79</td>
</tr>
<tr>
<td>4.7</td>
<td>Impact of vehicles’ density on end-to-end delay and overhead</td>
<td>80</td>
</tr>
<tr>
<td>4.8</td>
<td>Impact of traffic load on end-to-end delay and overhead</td>
<td>80</td>
</tr>
<tr>
<td>5.1</td>
<td>Basic Input-Output System</td>
<td>84</td>
</tr>
<tr>
<td>5.2</td>
<td>Abstract Scenario for multi-hop wireless network</td>
<td>90</td>
</tr>
<tr>
<td>5.3</td>
<td>LTE reference scenario</td>
<td>94</td>
</tr>
<tr>
<td>5.4</td>
<td>Bounding Function $g(x)$ for different $\theta$</td>
<td>97</td>
</tr>
<tr>
<td>5.5</td>
<td>Service Curve $\beta(x)$ for different $\theta$</td>
<td>97</td>
</tr>
<tr>
<td>5.6</td>
<td>Reference area for simulations and validation of model</td>
<td>100</td>
</tr>
<tr>
<td>5.7</td>
<td>Comparison of model bound and simulation results for data packets (10-nodes/group scenario)</td>
<td>102</td>
</tr>
<tr>
<td>5.8</td>
<td>Comparison of model bound and simulation results for Loc. Service packets (10-nodes/group scenario)</td>
<td>103</td>
</tr>
<tr>
<td>5.9</td>
<td>Comparison of model bound and simulation results for different hop count</td>
<td>103</td>
</tr>
<tr>
<td>5.10</td>
<td>Comparison of model bound and simulation results for different data rates</td>
<td>104</td>
</tr>
<tr>
<td>5.11</td>
<td>Comparison of model and simulation results for data packets (3-hop &amp; 20veh. scenario)</td>
<td>105</td>
</tr>
<tr>
<td>5.12</td>
<td>Comparison of end-to-end delay distribution for data traffic</td>
<td>105</td>
</tr>
<tr>
<td>5.13</td>
<td>Comparison of simulation results for signalling traffic</td>
<td>106</td>
</tr>
<tr>
<td>5.14</td>
<td>Comparison of throughput for different data rates</td>
<td>107</td>
</tr>
<tr>
<td>6.1</td>
<td>Original SCTP Markov Chain states with multihoming</td>
<td>110</td>
</tr>
<tr>
<td>6.2</td>
<td>Modified SCTP Markov Chain states with multihoming</td>
<td>112</td>
</tr>
<tr>
<td>6.3</td>
<td>Network model</td>
<td>113</td>
</tr>
</tbody>
</table>
6.4 Evaluation of RTT-aware SCTP model ........................................ 114
6.5 Path utilisation ratio ............................................................. 115
6.6 Provider’s revenue in flat pricing – opt-saturated and unsaturated regimes 118
6.7 Off-Loading Ratio $\kappa_{avg}$ for different traffic demands ................. 119
6.8 Algorithm for proposed CE-SCTP ........................................... 119
6.9 Hybrid Architecture for CE-SCTP ........................................... 121
6.10 Reference Scenario ............................................................ 122
6.11 Scenario with data flow 100kbps ......................................... 123
6.12 Scenario with data flow 150kbps ......................................... 124
6.13 Scenario with data flow 200kbps ......................................... 124
6.14 Traffic Distribution among users on the two networks ...................... 125
7.1 Visualised summary of contributions ....................................... 128
A.1 Performance comparison of CLWPR with topology based routing protocols in Scenario #1 ....................................................... 136
A.2 Performance evaluation of Carry-n-Forward mechanism in Scenario #1 137
List of Tables

2.1 ITS applications communication requirements ........................................ 17
2.2 Routing protocols taxonomy for VANETs ............................................. 19
2.3 Position-based routing protocols without navigation ............................... 21
2.4 Position-based routing protocols navigation .......................................... 22
2.5 Recovery methods for position-based routing protocols .......................... 26
2.6 Cross-Layer routing protocols ............................................................... 28
2.7 Summary of routing protocols for VANETs ............................................. 44
3.1 HELLO Message Information .............................................................. 49
3.2 Neighbor List for scenario in Figure 3.4 .................................................. 55
3.3 Scales of Pairwise Comparison .............................................................. 59
3.4 Simulation Parameters for CLWPR evaluation ....................................... 61
3.5 Optimal Values for CLWPR parameters ............................................... 67
4.1 Simulation Parameters for Location Service evaluation ........................... 77
5.1 SAC for different arrival types ............................................................... 85
5.2 Latency and rate terms of schedulers .................................................... 87
5.3 Standardized QCIs for LTE ................................................................. 96
5.4 Configuration Parameters for SNC evaluation ....................................... 101
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGF</td>
<td>Advanced Greedy Forwarding</td>
</tr>
<tr>
<td>AHP</td>
<td>Analytic Hierarchy Process</td>
</tr>
<tr>
<td>AP</td>
<td>Access Point</td>
</tr>
<tr>
<td>BS</td>
<td>Base Station</td>
</tr>
<tr>
<td>BTP</td>
<td>Basic Transport Protocol</td>
</tr>
<tr>
<td>CAR</td>
<td>Connectivity Aware Routing</td>
</tr>
<tr>
<td>CLWPR</td>
<td>Cross-Layer Weighted Position-based Routing</td>
</tr>
<tr>
<td>CMT</td>
<td>Concurrent Multipath Transfer</td>
</tr>
<tr>
<td>CSI</td>
<td>Channel State Information</td>
</tr>
<tr>
<td>DAB</td>
<td>Digital Audio Broadcasting</td>
</tr>
<tr>
<td>DSRC</td>
<td>Dedicated Short Range Communication</td>
</tr>
<tr>
<td>DVB</td>
<td>Digital Video Broadcasting</td>
</tr>
<tr>
<td>E2ED</td>
<td>End-to-End Delay</td>
</tr>
<tr>
<td>EDCA</td>
<td>Enhanced Distributed Channel Access</td>
</tr>
<tr>
<td>EPC</td>
<td>Evolved Packet Core</td>
</tr>
<tr>
<td>EPS</td>
<td>Evolved Packet System</td>
</tr>
<tr>
<td>ETSI</td>
<td>European Telecommunications Standards Institute</td>
</tr>
<tr>
<td>ETSI-GF</td>
<td>ETSI Greedy Forwarding</td>
</tr>
<tr>
<td>GN6ASL</td>
<td>GeoNetworking IPv6 Adaptation SubLayer</td>
</tr>
<tr>
<td>GPCR</td>
<td>Greedy Parameter Coordinator Routing</td>
</tr>
<tr>
<td>GPSR</td>
<td>Greedy Perimeter Stateless Routing</td>
</tr>
<tr>
<td>GyTAR</td>
<td>Greedy Traffic Aware Routing</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>ICT</td>
<td>Information and Communication Technologies</td>
</tr>
<tr>
<td>ITS</td>
<td>Intelligent Transport Systems</td>
</tr>
<tr>
<td>LOS</td>
<td>Line of Sight</td>
</tr>
<tr>
<td>LS</td>
<td>Location Services</td>
</tr>
<tr>
<td>LTE</td>
<td>Long Term Evolution</td>
</tr>
<tr>
<td>MANET</td>
<td>Mobile Ad-hoc Network</td>
</tr>
<tr>
<td>MBMS</td>
<td>Multimedia Broadcast Multicast Service</td>
</tr>
<tr>
<td>OBU</td>
<td>On-Board Unit</td>
</tr>
<tr>
<td>PDR</td>
<td>Packet Delivery Ratio</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>RSU</td>
<td>Road Side Unit</td>
</tr>
<tr>
<td>RTT</td>
<td>Round Trip Time</td>
</tr>
<tr>
<td>SAC</td>
<td>Stochastic Arrival Curve</td>
</tr>
<tr>
<td>SCTP</td>
<td>Stream Control Transport Protocol</td>
</tr>
<tr>
<td>SINR</td>
<td>Signal to Interference and Noise Ration</td>
</tr>
<tr>
<td>SNC</td>
<td>Stochastic Network Calculus</td>
</tr>
<tr>
<td>SSC</td>
<td>Stochastic Service Curve</td>
</tr>
<tr>
<td>V2I</td>
<td>Vehicle to Infrastructure</td>
</tr>
<tr>
<td>V2V</td>
<td>Vehicle to Vehicle</td>
</tr>
<tr>
<td>VANET</td>
<td>Vehicular Ad-hoc Network</td>
</tr>
<tr>
<td>WAVE</td>
<td>Wireless Access in Vehicular Environments</td>
</tr>
<tr>
<td>WINNER-II</td>
<td>Wireless World Initiative New Radio - Phase II</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Scope

People spent a significant amount of their time travelling in their vehicles. Based on a recent National Travel Survey \[1\] we are informed of the daily travel patterns for United Kingdom travellers, with an average trip length of 7 miles (\(\sim\)11km) and trip time of \(\sim\)1 hour/day. Moreover, the way we travel (Figure 1.1) and the reason we travel (Figure 1.2) play a key role on what is expected from our journeys as user experience. The current transportation systems are costly; according to a recent report from CISCO:

- **Global human costs** account for 8 million traffic accidents resulting in 1.3 million lives and 7 million injuries.

- **Global economical impacts** account for 90 billion hours spent in traffic jam, per year. To annual costs of vehicles in the US of 1.8 trillion and the infrastructure plus environmental costs of about 1.1 trillion (congestion accounts for more than 10\% of the costs), while in UK the average speed continued to decline to 24 miles/hr.

- **Environmental impacts** of transportation account for 220 million metric tons of CO2 produced each year. In UK, 26\% of green house gas (GHG) comes from Transport and 40\% of that from private cars and taxis.
Figure 1.1: How we travel? [1]

- Car: 65%
- Walk: 22%
- Bus: 6%
- Rail: 3%
- Cycle: 2%
- Other: 2%

Figure 1.2: Why we travel? [1]

- Commuting: 15%
- Entertainment: 15%
- Personal: 19%
- Shopping: 21%
- Business: 3%
- Education: 12%
- Visiting: 15%
1.2. Challenges and Motivations

Safety has always been one of the driving forces of innovation in automotive industry (Figure 1.3). From passive and active safety systems such as seatbelts and sensors, which alleviate the damage of an accident and alert drivers, we are currently at what is known as *connected mobility*. Intelligent Transportation Systems (ITS) are the building block for connected mobility. They are the outcome of coupling transportation systems with wireless communications offering users several innovative and effective services. It is envisaged that the future vehicles will be able to form ad-hoc networks among them, or connect to infrastructure in order to exchange important traffic and safety related information in highway and urban environments. Advances in wireless communications have made it possible to form Vehicular Ad-Hoc Networks (VANETs). Research on VANETs has gathered momentum due to the increased interest by the automotive industry and governments in an attempt to improve the quality and safety of future transport systems. Several ITS applications have been proposed with different communication requirements [2,3]. To this end, all aspects of reliable communication systems, including transport and routing protocols, have been considered to be optimised for efficient operation in such environments.

1.2 Challenges and Motivations

Although there are existing routing protocols for Mobile Ad-Hoc Networks (MANETs), importing them directly into VANETs exhibits unsatisfactory performance [5,6]. Some of the differences that distinguish VANETs from MANETs are the following. Unlike MANETs that usually rely on batteries with restricted energy, VANETs lack strict energy constraints due to the power supply provided by a vehicle’s engine. The computational power and storage capacity are also larger than other MANETs, which enables developers to design more complex systems. A key characteristic of VANETs is the mobility of the nodes. Depending on the environment, one can have high velocities in straight roads like in highway environments, or lower speeds with frequent stops and turnings when moving inside cities. Contrary to random models that are used to describe the movement of MANET nodes, vehicles are restricted by the underlying road network, which makes it possible to predict their positions. Moreover, the environment
Figure 1.3: Impact levels of Intelligent Transport Systems
characteristics such as buildings and other vehicles oppose challenging conditions for the wireless communications. Finally, the density of nodes varies over time which affects the connectivity of the network. These characteristics impose several challenges to the design of reliable communications:

- **Connectivity Management**
  The network topology is very dynamic and the environment is affecting the communication channel dramatically, both in urban areas as well as highways. This poses significant challenges in order to establish a reliable end-to-end connection. Most of the recently proposed routing protocols for VANETs are based on the position of the nodes, rather than the network topology, to make their forwarding decisions. Although the mobility of vehicles is relatively high, it is not random. Therefore, using the appropriate mechanisms it can be predicted. On the other hand, this information should be accessible by all nodes. Maintaining real-time position information and making it available to any node that requests it with minimum overhead to the network, is challenging. Due to the high mobility and the environmental characteristics, the communication channel is difficult to be modelled. Special care has to be taken when considering line-of-sight and non-line-of-sight components of wireless communications. In addition, interference caused by other communicating vehicles is high which makes even more difficult to maintain a communication link.

- **Quality of Service Provision**
  Different ITS applications require different levels of Quality of Service (QoS). For example, safety applications need fast real-time data dissemination (e.g. 100ms latency), whereas infotainment services are usually best effort but with higher data rate requirements. Ensuring that packets are forwarded through the appropriate nodes that minimizes delays and increases the probability of reception is not trivial. This may contradict with the Network Neutrality that Federal Communications Commission (FCC) recently ruled in favour.

- **Heterogeneous Network Support**
  Typically, vehicles use only one network interface to connect to other vehicles or
to infrastructure. However, the proliferation of multiple network interfaces, e.g. dedicated short-range communications (DSRC) and cellular (3GPP LTE), has spurred the development of multi-homing and multi-path technologies in vehicular communications. Such systems are an interesting research field, aiming to bridge the gap between intermittent, low capacity and low latency DSRC systems, and ubiquitous coverage, high capacity and higher latency cellular systems. They are also used for off-loading and increase reliability of vehicular communications.

- **Security and Privacy Provision**

  All these network interfaces available for on-board and off-board communications leave open doors to attackers that could potentially take over the control of the vehicle. Further, privacy is a major subject in vehicular communication systems. Vehicle details such as licence plate number, location, driving patterns could be illegitimately exploited by the adversaries, companies, and government. Privacy refers to the capability that ensures personal identifying information of the drivers is not disclosed to third parties. Hence, the network often and by default is required to ensure that activities of drivers cannot be traced by the adversaries. However, conditional privacy stipulates that it should be possible to reveal the identity of offending drivers for revocation and criminal prosecution. Anonymity is a technique of hiding the physical identity of a vehicle such as IP address and electronic number plates. The Vehicular Public Key Infrastructure (VPKI) should ensure a means to offer conditional anonymity to drivers by separating vehicular information from personal information about drivers.

A reference architecture of communication protocols stack for vehicular communications is specified for DRIVE C2X in [7] as depicted in Figure 1.4a. It is adopted from the ETSI specifications [8]. The work of this thesis tackles the connectivity management and heterogeneous network support research challenges by providing solutions fit for the ITS communication protocol stack in Figure 1.4a. Particularly, it relates to the Facility, Network and Access layers of this architecture, combining them in a cross-layer architecture with interfaces through the ITS Management. Thus providing enhancements for each layer of the ITS protocol stack (excluding security) to increase
1.2. Challenges and Motivations

Efficiency and reliability of vehicular communications. Access layer is usually based on the PHY and MAC specifications of the IEEE 802.11p standard for ad-hoc communication. However, 3G or 4G technologies can also be used for applications, management, and testing purposes communicating with infrastructure. Additionally, GPS and Sensor equipment are used to provide position information and interface to real world. Networking and Transport layers handle reliable delivery of messages across the network and play key roles in communications. These layers implement a number of functionalities as shown in Figure [1.4b]. UDP and TCP can be used as transport layer protocols. Basic Transport Protocol (BTP) [9], a UDP-like transport protocol, is commonly used with geographic routing. In addition, ITS specific transport protocols have been designed to cope with the characteristics of vehicular traffic, e.g., Vehicular Transport Protocol (VTP) [10] and Vehicular Information Transfer Protocol (VITP) [11]. Transport protocols are further reviewed in subsection 2.3.3. The network layer implements functionalities such as routing, addressing, mobility management, and other. Position-based routing is usually used for ad-hoc communications among vehicles and Road Side Units (RSUs) and relies on a geo-addressing scheme, which is specified in [12]. In addition, geo-addressing requires a “translation” mechanism from IP addresses to geo-networking addresses. This is usually done by a GeoNetworking IPv6 Adaptation SubLayer (GN6ASL) [13]. In position-based routing protocols, each node maintains a local database which stores the position of its neighbours that is learnt through the neighbour discovery mechanism. Moreover, position-based routing relies on a location service in order to identify the position of a destination node. Each node can determine its own location and get navigation information from the facility layer in Figure [1.4a].

Finally, an important part of system’s architecture, that spans alongside all layers, is Security. The requirements set for security include aspects such as data integrity, authentication, and privacy, as well as detection and resilience against attacks. The security mechanisms for the ETSI system architecture are described in [14], however it is outside of the scope of this thesis and thus it is not further discussed.

There has been a number of recent position-based unicast routing proposals for VANETs in the literature (please refer to Section 2.3.1) that can be categorised as: 1) greedy forwarding; 2) mobility assisted; 3) cross-layered. However, simple greedy approaches
Chapter 1. Introduction

(a) ITS communication protocol stack

(b) Network and Access Layer

Figure 1.4: ITS reference architecture (a), Network and Access Layers (b)

do not perform well in dynamic networks such as VANETs, due to the effect of local maximum problem. Although algorithms that take into account mobility information improve the performance of the protocol in terms of received packets, they increase network overhead due to additional information exchange needed. In addition, it is known that exploitation of cross layer information can benefit performance of routing protocols. To the best of our knowledge, the existing proposals do not consider the aforementioned factors within a coherent framework, nor investigate the effects of communication and environment parameters on forwarding decisions. Moreover, for the proper functionality of those position-based routing protocols, an efficient location service has to operate. It has to provide location information to vehicles requesting the position of another vehicle they want to forward a packet to. The rapid increase of multi-homed devices on vehicles with both DSRC and LTE connectivity, gives an opportunity to split data and signalling traffic on the two networks, in order to reduce the latency for both data and signalling traffic. Finally, the use of cellular traffic comes with higher cost, compared to none or very small for WiFi usage. Therefore, the selection of alternate interface in a multi-homed device should take into account the pricing schemes employed in order to have a cost-efficient as well as reliable communication.

Motivated by this demand, this work tries to give an insight on how we can provide reliable wireless communications in such a demanding environment. We target efficient position-based routing protocols for VANETs, location managing services and multi-homing support in transport layer with cost-aware interface selection. Specifically, we
1.3 Objectives

seek to provide answers to the following research questions (RQ):

*RQ1* Which node should be selected as next-hop for efficient and reliable unicast data dissemination? How can we exploit mobility and environment characteristics to improve routing decisions?

*RQ2* How to efficiently deliver position information needed for routing? Should the same communication channel be used for data and signalling dissemination?

*RQ3* Which interface should be used for cost-efficient data transfers? How could end-users’ and provider’s economic benefits increase with exploitation of multi-homing support?

1.3 Objectives

The European Telecommunications Standards Institute (ETSI) has defined several objectives for the performance of network and transport layer functionality in ITS networks [16]. As the proposed works should fit in ETSI’s protocol stack architecture, specifically in the ITS network and transport layer depicted in Figure 1.4, they should tackle the same objectives. In general, the ITS network and transport layer shall provide low-latency communications and reliable communications with the highest reliability for safety messages. In addition, it shall keep signalling, routing and packet forwarding overhead low. Further, it shall be fair among different nodes with respect to bandwidth usage considering the type of messages and robust against security attacks, and malfunctions in ITS stations. Finally, it shall be able to work in scenarios with low and high density of GeoNetworking-enabled nodes. These objectives also underlie this work and are further specified to answer the three research questions as follows.

Routing plays a major role in any network. It provides the service of forwarding packets towards other nodes seamlessly to the user. Specially in VANETs, due to the characteristics described above, routing is faced with great challenges to find the best possible path for a packet. In order to ensure the highest reliability in the design of the forwarding mechanism, QoS requirements from the application should be met.
Therefore, we consider a cross-layer approach by employing PHY and MAC layer at the routing decision. The objective is to increase the performance of the routing protocol by intelligently selecting the next forwarding node. This can be measured with indicators such as packet delivery ratio, end-to-end delay and routing overhead.

Furthermore, and since the proposed routing protocol is based on node positioning, a proper mechanism to maintain and provide the position of any node has to be designed. Several Location Services have been proposed but few of them are designed for vehicular environments. Therefore, a new architectural design has been proposed aiming at providing accurate position information with the least cost of overhead and delay, benefiting from distributed database architecture and node co-operation. By utilizing multiple network interfaces, we split data traffic on DSRC network and signalling for the location service on LTE network.

Finally, different networks may provide variable end-to-end path quality and incur in significant cost dissimilarity. Hence, proper care should be taken in using multi-homing support from the transport layer protocol. A synergy of QoS and cost metrics should be employed in the selection of the best path.

1.4 Contributions

The contributions of this work are grouped in three categories, each answering one of the Research Questions stated previously.

Routing Protocol: An innovative and effective position based routing protocol is proposed that takes into account all the major network and environment parameters from PHY, MAC, and network layers. This work considers node reliability and the effect of carry-n-forwarded messages, indicators not previously considered by other works and characteristics that are significant in highly dynamic VANETs. Furthermore, we propose an adaptive HELLO broadcasting scheme to cope with communication overhead; another important issue for VANETs. Analytic Hierarchy Process (AHP) is employed to optimally combine multiple decision criteria involved in a fast forwarding mechanism, which results in both qualitatively and
quantitative findings for the effects of mobility, link quality and node utilisation related information in forwarding decisions. Comprehensive performance analysis in representative urban scenarios are performed that take into account realistic propagation models and real traffic. To the best of our knowledge, this is the first work that examines the effects of multiple parameters (communication and environment related) on the performance of routing in VANETs in a systematic framework.

**Location Service:** We propose and evaluate a centralised location service architecture that is based on the separation of data and signalling on different networks. In order to off-load the wireless 802.11p-based access network, we propose a hybrid solution that utilises also existing cellular network (e.g. LTE). Furthermore, an innovative approach based on stochastic network calculus (SNC) methodology is introduced to find the upper bounds on end-to-end delay in vehicular networks employing position-based routing. The proposed analytical models are validated using realistic simulations in NS3 environment. We compared three networking architectures based on: (a) short-range ad-hoc communications only, (b) cellular communication only, and (c) hybrid ad-hoc/cellular communications. The results of the performance evaluation suggest that in higher traffic loads and higher vehicle densities, homogeneous networks (e.g. 802.11p, LTE) suffer from congestion. The proposed heterogeneous architecture however, can cope better in such scenarios.

**Transport Protocol:** Multi-homing support is a feature of SCTP protocol, however by default it is only used when the primary interface becomes unreachable. Proposals to dynamically use all available interfaces have been analysed only through simulations. In this work, an analytical model for SCTP protocol is proposed where the interface with minimum RTT is dynamically selected. Moreover, we propose a novel hybrid vehicular networking architecture enabling cost-efficient multi-homing transport layer. Then, a cost-aware SCTP scheme for the proposed hybrid architecture is proposed, which benefits both the user and the network provider.
A complete list of the publications resulted during this research is presented in Appendix B. The source code for CLWPR is publicly available under GPLv2 Licence and has been submitted for review by the NS-3 community along with other minor patches to the NS-3 code.

1.5 Thesis Overview

The remainder of this thesis is organised in six chapters as follows:

Chapter 2 gives an overview of intelligent transport systems and the applications realised in these systems. Furthermore, related works on vehicular communications are reviewed. In particular, routing protocols, location services, transport protocols, propagation models, and delay modelling approaches are examined.

Chapter 3 presents the proposed routing protocol, CLWPR. This includes the protocol design and methodology of parameter adjustment. In addition, extensive performance evaluation of the proposed routing protocol and comparison with related protocols is conducted in a NS-3 based simulation environment.

Chapter 4 introduces a hybrid network architecture comprising short range ad-hoc network based on IEEE 802.11p and long range cellular network based on 3GPP LTE. The proposed network architecture is used to realise a centralised location service, which is evaluated through extensive simulations.

Chapter 5 presents the proposed model for end-to-end upper bound on delay using Stochastic Network Calculus. Three architectures are evaluated and validated through simulations; (a) only a vehicular ad-hoc network based on a short range communication technology, (b) only a cellular network with a larger coverage area, and (c) a hybrid network comprising an ad-hoc and a cellular network as introduced in previous chapter.

1 http://personal.ee.surrey.ac.uk/Personal/K.Katsaros/ns3.html
2 The source code is submitted for review at http://codereview.appspot.com/5343044
Chapter 6 introduces an analytical model for a utility-based SCTP multi-homing, which selects the primary interface based on minimum RTT. Furthermore, a proposal for a hybrid vehicular networking architecture is presented which will facilitate a cost-effective SCTP protocol. This is later evaluated in the hybrid vehicular network.

Chapter 7 summarises and concludes this thesis providing ideas future expansion and open research challenges.
Chapter 2

Background and Related Works

This chapter gives an in depth analysis of the architecture and system model of Intelligent Transport Systems and examines related work on vehicular communications. The structure of this chapter is as follows. Sections 2.1 and 2.2 give an overview of ITS and the applications realised in ITS with particular communication requirements. Section 2.3 presents related work on vehicular communications with respect to routing protocols, location services, transport protocols, propagation models and delay models. Finally, section 2.4 summarises the findings.

2.1 Intelligent Transport Systems

Rapid increase in the number of vehicles in the recent decades has resulted in growing concerns about the adverse environmental impacts and safety issues in land transport systems [17]. Thus, Intelligent Transportation Systems have emerged as promising solutions for the future effective and environment friendly transport systems. ITS aim to apply Information and Communication Technologies (ICT) to improve safety and efficiency as well as the passenger experience in modern transport systems. Figure 2.1 demonstrates potential uses cases of ITS, that can span from adaptive cruise control to passenger information and from short range vehicle-to-vehicle to satellite communications.
For each of the use case, different access technology can be used, including satellite, terrestrial broadcast, mobile and dedicated short-range communications. Governmental and standardization bodies in the U.S. (National Highway Traffic Safety Administration, IEEE, CEN), Europe (European Commission, C2C-CC, ETSI) and Asia (Ministry of Land, Infrastructure, Transport and Tourism of Japan, China Communications Standards Association (CCSA)), are in the phase of standardising and regulating the different technologies that will facilitate inter-vehicle and vehicle-to-infrastructure communications \[18\]–\[21\]. In addition, several projects are conducting research with field operation tests in order to assess the impacts of ITS applications with respect to safety, traffic efficiency and environmental impacts, e.g. DRIVE-C2X \[22\], euroFOT \[23\], eCoMove \[24\], SHRP-2 \[25\].

### 2.2 Applications for ITS

ITS applications can be categorised in three basic groups depending on their functionality; safety, efficiency and comfort. Each of these categories has different communication
2.3. Related Works

2.3.1 Routing Protocols

In the subsequent sections we investigate the different methods of forwarding a packet through a wireless ad-hoc network, focusing on position-based mechanisms, which are more suitable for VANETs. We expand on different forwarding techniques specifically those that employ navigation information. Providing ITS applications with a certain QoS is a challenge, which cross-layer network protocols we review below aim to tackle.

---

**Table 2.1: ITS applications communication requirements**

<table>
<thead>
<tr>
<th></th>
<th>Safety</th>
<th>Efficiency</th>
<th>Comfort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Collision</td>
<td>Road Sign</td>
<td>Traffic Management</td>
</tr>
<tr>
<td>Com. Type</td>
<td>Avoidance</td>
<td>Management</td>
<td></td>
</tr>
<tr>
<td>Data Rate</td>
<td>V2V / V2I</td>
<td>V2I</td>
<td>I2V</td>
</tr>
<tr>
<td></td>
<td>up / down</td>
<td>up / down</td>
<td>down</td>
</tr>
<tr>
<td></td>
<td>Very low</td>
<td>low</td>
<td>low</td>
</tr>
<tr>
<td></td>
<td>medium / medium</td>
<td>medium / med</td>
<td>medium / high</td>
</tr>
<tr>
<td></td>
<td>uni / broadcast</td>
<td>uni / broadcast</td>
<td>uni / broadcast</td>
</tr>
<tr>
<td>Reliability</td>
<td>high</td>
<td>high</td>
<td>medium / high</td>
</tr>
<tr>
<td>Priority</td>
<td>high</td>
<td>high</td>
<td>medium</td>
</tr>
<tr>
<td>Supported</td>
<td>DSRC</td>
<td>DSRC</td>
<td>DSRC</td>
</tr>
<tr>
<td>Technology</td>
<td></td>
<td></td>
<td>DSRC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In general, safety related applications have low latency and high priority requirements, while their locality is restricted to vehicles and roadside units in the vicinity of the event. Dedicated short range communications (DSRC\footnote{IEEE WAVE in U.S. and ETSI ITS G5 in Europe}) is proposed as the main technology that can support these applications with inherent broadcast and priority mechanism. On the other hand, comfort applications require longer range communications with higher data rates and have higher tolerance on latencies. Such applications can be served also by cellular or other wide area access technologies, in addition to DSRC. This type of applications, which require one-to-one (unicast) communication, is the focus of this research work.
2.3.1.1 Routing protocols taxonomy for VANETs

Routing protocols can be categorised according to their design as topology-based, hierarchical (clustering), flooding (broadcasting), and geographical (position-based) as presented in Table 2.2. Topology-based routing rely on the network graph that is composed by the nodes and the communication links between them. They are divided into proactive (table driven), such as OLSR [28] and DSDV [29], and reactive (on demand) protocols, such as AODV [30] and DSR [31]. Proactive protocols introduce network overhead which increases as the size of the network topology is increased in order to keep their routing tables updated. On the other hand, reactive protocols add a delay in the beginning of the communication in order to discover a route whilst flooding the network with this query. Furthermore, the dynamic topology of a vehicular network will soon make the former route obsolete and thus a new query will be needed. There are also hybrid protocols, such as TORA [32] and ZPR [33], which combine characteristics of proactive and on demand protocols. Hierarchical protocols, such as HRS [34], divide the network into clusters, which share some common characteristics for a period of time. The inter-cluster communication is achieved through specified nodes which act as gateways. The aim of these protocols is the optimisation of resource allocation but the dynamics of vehicular networks impose frequent changes on the clustering formation which in turn increases the overhead needed to maintain a cluster. The simplest way of disseminating a packet is to flood it in the network. This way, the complexity of the routing protocol is minimised but the overhead is exponentially increased. In order to use this kind of protocols in VANETs, several optimisations have been proposed to reduce the number of re-broadcast packets but still the bandwidth is unfairly used.

The last category of routing protocols, geographical, is the one which best fits vehicular ad-hoc networks. The principle behind geographic routing is that forwarding decisions are made based on the position of the nodes and not on the network graph. Two fundamental assumptions are made in these protocols. First, that a node is able to know its own position. Such an assumption is valid since the use of GPS technology is widespread and every vehicle can be equipped with such a device. Apart from GPS, other means of positioning have been developed that can be used, such as triangulation.
Table 2.2: Routing protocols taxonomy for VANETs

<table>
<thead>
<tr>
<th>Topology Based</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proactive (e.g. OLSR [28])</td>
<td>Do not flood entire network, Fast path selection</td>
<td>Overhead to maintain tables</td>
</tr>
<tr>
<td>Reactive (e.g. AODV [30])</td>
<td>Do not maintain routing tables</td>
<td>Initial delay for route discovery, Flood a route request</td>
</tr>
<tr>
<td>Hybrid (e.g. TORA [32])</td>
<td>Combination of proactive and reactive in different operation stages</td>
<td></td>
</tr>
<tr>
<td>Hierarchical (e.g. HRS [34])</td>
<td>Exploit clusters with similar characteristics</td>
<td>Overhead to maintain clusters</td>
</tr>
<tr>
<td>Flooding (e.g. Epidemic [36])</td>
<td>Low complexity, high data reception ratio</td>
<td>Flood entire network</td>
</tr>
<tr>
<td>Position Based</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without Navigation (e.g. GPSR [37])</td>
<td>Rely on local information only</td>
<td>Need a location service, more prone to local maximum problem</td>
</tr>
<tr>
<td>With Navigation (e.g. GyTAR [38])</td>
<td>Exploit mobility of nodes, less prone to local maximum problem</td>
<td>Need a location service, increase in overhead due to enhanced beaconting</td>
</tr>
</tbody>
</table>

The second assumption, and most significant, is that every node knows or is able to know the position of the destination when it is needed. This is achieved with two methods; either the use of location services such as HLS [35] that manages the position of all nodes or by broadcasting a query for the position of the node like on-demand protocols do when they initiate a communication. The characteristics that favour geographical routing protocols in VANETs over the rest are the fact that they scale better in large networks since they only use localised information (only neighbouring information) to select the next forwarding node instead of the complete network graph that topology protocols need. Also, the routing overhead is less than flooding protocols since they only broadcast 1-hop beacon messages as a mean of neighbour discovery. Finally, compared with hierarchical protocols, geographical do not have the clustering overhead. Thus, the use of geographical routing is vital in VANETs due to the highly dynamic topologies and the potential large number of nodes.

2.3.1.2 Geographic Forwarding Mechanisms

Position-based or geographic routing protocols have emerged as promising solution for routing in VANETs. Geographic routing protocols were initially introduced in the 1980s [39, 40], but they were not well received at the time due to high cost and
inaccuracy of positioning devices. However, with proliferation of cheap and accurate position systems, such as GPS, position based routing became popular once again in the recent years and is set as standard in ETSI GeoNetworking [12]. These protocols employ some aspects of proactive routing protocols, where periodic broadcast messages are used for neighbour discovery (neighbour is a node that can be directly communicated), and some aspects of reactive routing protocols, for discovering the geographical location of the destination nodes, using some sort of location services.

The first protocols that were designed for MANETs assumed that nodes are randomly distributed and their mobility is relatively low. However, in VANETs, nodes travel on roads and navigation systems can provide additional information, which could be used for routing. Therefore, it is appropriate to distinguish the forwarding mechanisms into two categories; those using only positioning information and those employing navigation as well. Finally, we define the local maximum problem that geographic routing protocols are faced with and provide a summary of recovery mechanisms used in VANETs to overcome this problem. A recent survey of position-based routing protocols for VANETs can be found in [41].

**Forwarding without Navigation:** In this section, we focus on unicast ad-hoc protocols using position information only, a legacy from MANETs, and more specifically how a node selects the next forwarding node based solely on geographically related information. A list of them is given in Table 2.3 with the corresponding routing metric. We start with what is known as Greedy Forwarding (GF) [40]. With this method the next forwarding node is selected based on the geographic (Euclidean) distance from the destination. As shown in the example scenario of Fig. 2.2 in GF policy, source Node S will forward its packets towards Node #4, which is the closest node to the destination Node D. This policy is employed in several protocols such as GPSR [37], ETSI-GF [12] and Finn et al. [40]. A different approach is to take the “Most Forward within Radius” (MFR), which is proposed in [39]. This scheme suggests that the node to be selected will provide the most forwarding distance on the direct line from the source towards the destination. This can be calculated using the cosine of the angle that is formed from a node, the source and the destination. In our example, \( \cos \hat{S}D \) provides the greatest
2.3. Related Works

Table 2.3: Position-based routing protocols without navigation

<table>
<thead>
<tr>
<th>Routing Protocol</th>
<th>Routing Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPSR [37]</td>
<td>euclidean distance</td>
</tr>
<tr>
<td>ETSI-GF [12]</td>
<td></td>
</tr>
<tr>
<td>Finn et al. [40]</td>
<td></td>
</tr>
<tr>
<td>CGGC [46]</td>
<td></td>
</tr>
<tr>
<td>MoVe [47]</td>
<td></td>
</tr>
<tr>
<td>SAR [48]</td>
<td></td>
</tr>
<tr>
<td>Gasari et al. [49]</td>
<td></td>
</tr>
<tr>
<td>Basagni et al. [50]</td>
<td></td>
</tr>
<tr>
<td>Improved GPSR [51]</td>
<td></td>
</tr>
<tr>
<td>LTR [52]</td>
<td></td>
</tr>
<tr>
<td>Yang et al. [53]</td>
<td></td>
</tr>
<tr>
<td>Chen et al. [54]</td>
<td></td>
</tr>
<tr>
<td>MFR [39]</td>
<td>most forward radius</td>
</tr>
<tr>
<td>NFP [42]</td>
<td>nearest forwarding progress</td>
</tr>
<tr>
<td>Nelson et al. [43]</td>
<td>random positive progress</td>
</tr>
<tr>
<td>Compass [44]</td>
<td>angle</td>
</tr>
<tr>
<td>PROMPT [55]</td>
<td>distance plus MAC statistics</td>
</tr>
<tr>
<td>Barghi et al. [56]</td>
<td>distance plus mobility metrics</td>
</tr>
<tr>
<td>VTP [10]</td>
<td>distance plus bandwidth availability</td>
</tr>
<tr>
<td>Zhou et al. [57]</td>
<td>distance plus rate and MAC info</td>
</tr>
</tbody>
</table>

progress towards the destination and thus Node #1 is selected. On the other hand, the “Nearest Forwarding Progress” (NFP) scheme was proposed in [42] which selected the node with the least progress (Node #3 from the example). This is proposed in order to minimise transmission power so that interference and power consumption are reduced. The third approach that uses the notion of progress was made in [43], which proposes to randomly select one of the nodes that provide a positive progress towards the destination (any of the nodes #1 - #4 from Fig. 2.2). The last greedy approach, known as compass routing [44], tries to minimise the angle of the selected node and the direct line between source and destination. In our example, using this method, Node #2 would be selected because the angle $\hat{\angle}SD$ is the smallest. All these approaches are based on random mobility model (such as Random Waypoint), where each node (vehicle) can take any direction/speed. Such models are not suitable for VANETs with the constraints of the roads. More fitting mobility approaches for urban environments, which describe the movement of vehicles in cities more realistically, are the Manhattan Grid or real road networks [45].
Table 2.4: Position-based routing protocols navigation

<table>
<thead>
<tr>
<th>Routing Protocol</th>
<th>Routing Metric</th>
<th>Navigation Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAR [58], GPSRJ+ [59]</td>
<td>distance (euclidean)</td>
<td>number of intersections</td>
</tr>
<tr>
<td>GPCR [60]</td>
<td>distance (euclidean)</td>
<td>number of intersections</td>
</tr>
<tr>
<td>GyTAR [38]</td>
<td>distance (curve-metric)</td>
<td>traffic density and intersections</td>
</tr>
<tr>
<td>VADD [61]</td>
<td>distance (euclidean)</td>
<td>prediction, road speed limit</td>
</tr>
<tr>
<td>A-STAR [62]</td>
<td>distance (euclidean)</td>
<td>position prediction</td>
</tr>
<tr>
<td>AGF-GPSR [63]</td>
<td>distance (euclidean)</td>
<td>number of intersections</td>
</tr>
<tr>
<td>Optimized GPSR [64]</td>
<td>distance (euclidean)</td>
<td>number of intersections</td>
</tr>
<tr>
<td>MP2R [65], MAGF [66]</td>
<td>distance (euclidean)</td>
<td>prediction</td>
</tr>
<tr>
<td>GPSR-L [67]</td>
<td>distance (euclidean)</td>
<td>lifetime</td>
</tr>
</tbody>
</table>

Forwarding with Navigation: To increase the performance of routing in VANETs, protocols which employ navigation information are introduced. The knowledge of the underlying road topology can be of great importance to improve the design of a routing protocol. Table 2.4 lists the reviewed routing protocols that exploit some sort of navigation information; mainly regarding intersections. Using Fig. 2.3 as a reference for this part of the report we analyse the different schemes that are proposed. It has to be mentioned that in addition to the two previous basic assumptions (use of position system and location service), a third assumption has to be made for this kind of schemes. Nodes should be aware of the road network which again is a valid assumption since most of the vehicles are equipped with navigation devices that can provide such functionality.

Schemes such as Advanced Greedy (AG) [58] and Restricted Greedy (RG) [68] define “anchor” points at each intersection (e.g. I-1, I-2, I-3, and I-4 in Fig. 2.3). A node will search for a route towards the destination within the graph of interconnected junctions.
using a well-known algorithm, such as Dijkstra, and identify the minimum number of
intersections that a packet has to pass through. Then, the node will try to forward
the packet towards the first intersection using one of the previous map-less greedy
approaches. Once the packet has reached a node at the intersection (e.g. node #1) it
will then be forwarded towards the next intersection node using again a greedy method.

Other protocols that use this kind of approach are BAHG [69], CAR [58], GPCR [60]
and GyTAR [38]. One optimisation of this approach is made in GPSRJ+ [59] where
the forwarding node can predict the road that the packet will follow and thus skip
the intersection (e.g. forward to node #4 or #2 directly instead of #1). Therefore, a
decrease in the number of hops will be made. The beacons that each node broadcasts, as
means of neighbour discovery, could not only include their position but also their speed,
heading etc. Using this additional information, a node can make smarter decisions on
the forwarding nodes (e.g. forward towards nodes travelling on the same direction).

Protocols that use this scheme include VADD [61], A-STAR [62], AGF-GPSR [63] and
Optimized GPSR [64]. Similar to the latter, using the information about velocity, a
node can predict the current position of another node from its latest known position
and the time difference between present time and the time it received the beacon. This
method is used in VADD [61], MP2R [65], and MAGF [66]. Using prediction, a node
can make more accurate decisions regarding forwarding, thus there is an increase in the
performance of the routing protocol. GPSR-L [67] introduced the concept of lifetime
of a communication link in the routing. Using the information about the speed and
position of a node, it can predict the time it will remain in the communication range
and thus select the forwarding node accordingly. Finally, more advanced schemes use
information about the vehicle traffic and the road network such as the maximum speed
of a road (VADD [61]) and the traffic density (GyTAR [38]). The disadvantage of
“anchor” approaches is that they are not very dynamic. If the destination changes
its position, the optimal sequence of intersections should be re-calculated. Also, the
overhead is increased since this sequence of intersections is included in each packet.

On the other hand though, prediction is a mechanism that can potentially increase the
performance of the protocol that employs it. The use of traffic information, although it
increases the probability of finding a better forwarding node, increases also the overhead
introduced by the protocol, which is undesirable in large ad-hoc networks such as VANETs.

**Recovery Strategies:** In the previous sections we presented various greedy forwarding methods used in routing protocols for VANETs. The improvements that are proposed using navigation and other information aim at minimising the probability that a node will fall into the *local maximum* state. A node falls in the state of *local maximum* when it is the closest one towards the destination without being inside the communication range of the destination as shown in Figure 2.4. In this example, the Greedy Forwarding policy is employed. The same problem can be observed with any other forwarding policy. The incision between the circle with radius the communication range of a node and the circle with radius the distance between source and destination (grey area) does not include any other node. However, there might be other possible routes (S-1-6-7-D and S-5-9-8-D) to reach the destination, which can be found by employing a recovery strategy (Table 2.5). The first and simplest approach used when a node falls in the local maximum state, is to drop the packet. Although such an approach
may seem easy, it produces high packet loss rate and therefore is not suggested. One solution is the Enhanced Greedy forwarding used in CGGC protocol \[46\]. This proposes to delay the packet for a short period (random) and then try to resend it hoping that the node has left the local maximum state. If again there is no node to forward it to, the packet is dropped. This approach is improved by actively selecting when a packet that will be resent is known as \textit{carry-n-forward} or “mule” \[70\]. Each node has a buffer with either limited time or limited size that stores packets which could not be forwarded with greedy methods and have fallen into \textit{local maximum} state. Packets will then be transferred using vehicle’s speed until another forwarding node can be found. If the packets time out or there is not enough space in the buffer, they are dropped. Various protocols use this scheme, e.g. SAR \[48\], MoVe \[47\], VADD \[61\], and GyTAR \[71\]. A colouring mechanism is used instead of the \textit{carry-n-forward} in some MANET routing protocols \[49\], \[50\]. A node that is in the local maximum state changes its “colour” and packets are forwarded towards “greener” nodes. As we discussed previously, schemes such as AG and RG define anchor points that a packet has to be forwarded through. If a node falls in the local maximum state, one solution is to try to re-route the packet through different intersections, like in the colouring mechanism. Improving this approach by deleting the road segment that the local maximum appeared and then re-routing was proposed by A-STAR \[62\] and improved GPSR \[51\].

Using the example in Figure 2.3, if the source had selected the route I-1:I-2:I-4 to forward its packets, it would be faced with local maximum at intersection I-2. Adopting the latter method, the road segment between I-2 and I-4 would be deleted from the network and thus the only way to forward the packets would be through the route I-1:I-3:I-4. A similar approach is proposed in Depth-First-Search (DFS) \[72\] where the node memorises and deletes the paths that are defective. In GPCR \[68\] there are two kinds of nodes, coordinators nodes located at junctions and simple nodes. If a coordinator node is faced with the local maximum problem, then the right hand rule is used to select on which road segment to forward the packet assuming that the topology if the city is a planar graph. It is not mentioned if there is a recovery strategy for simple nodes. Finally, a more complex recovery strategy known as perimeter routing was proposed in GPSR \[37\]. It suggests that a node can generate the planar graph of the network
Chapter 2. Background and Related Works

Figure 2.4: Local Maximum Example

and from that using the right hand rule can find a path towards the destination. Using the reference network in Figure 2.4 node S would have to forward the packet to node #1 even though that is not the closest one to the destination. Such approach is not useful in VANETs due to the mobility and environment constraints. However, caching packets that do not have strict QoS requirement is a logical approach since there are not strict limitations on memory for VANETs and mobility is relatively high.

Table 2.5: Recovery methods for position-based routing protocols

<table>
<thead>
<tr>
<th>Routing Protocol</th>
<th>Recovery Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPSR [37]</td>
<td>Perimeter routing</td>
</tr>
<tr>
<td>GPSR-L [67]</td>
<td></td>
</tr>
<tr>
<td>GPCR [60]</td>
<td>Right hand rule</td>
</tr>
<tr>
<td>GyTAR [38]</td>
<td>Carry-n-forward</td>
</tr>
<tr>
<td>VADD [61]</td>
<td></td>
</tr>
<tr>
<td>MoVe [47]</td>
<td></td>
</tr>
<tr>
<td>SAR [48]</td>
<td></td>
</tr>
<tr>
<td>A-STAR [62]</td>
<td>Reroute with new anchor points</td>
</tr>
<tr>
<td>AGF-GPSR [63]</td>
<td></td>
</tr>
<tr>
<td>Improved GPSR [51]</td>
<td></td>
</tr>
<tr>
<td>CGGC [46]</td>
<td>Random delay on packet re-transmission</td>
</tr>
</tbody>
</table>
2.3. Related Works

2.3.1.3 Cross-Layering Designs

The development of communications is primarily based on the seven layered OSI model where layers are able to interact only with the adjacent layers. The introduction of internet has reformed this 7-layered approach to a compact 5-layered that is mostly known today. Each of these layers has a distinct functionality. With respect to wireless ad-hoc networks these functionalities can be described as follows. Adopting a bottom up approach, the physical layer (PHY) is responsible to make the actual data transmission, adapting to the changes of the wireless link and perform all the signal processing mechanisms. Above that, there is the data link layer with the Media Access Control (MAC) sub-layer. They are responsible to minimise collisions on the shared channel, provide fairness among the users and reliability by detecting errors from the PHY. In the middle lays the Network layer (NET) which maintains seamless connectivity among the nodes using a routing protocol. It is also in charge of distributing the information about the communication link used to maintain this connectivity. On top of NET, is the Transport layer which provides transparent transfer of data between end users with reliability, or not, depending on the used protocol. The most commonly used protocols are TCP (reliable) and UDP (unreliable). At the top of the 5-layered model stands the Application layer (APP) which runs the user application. We can see distinct functionalities for each one of these layers. However, in order to support adaptability to the challenging vehicular environment and perform certain performance optimisations, a new approach of interconnected layers was proposed, known as cross-layering [73,74].

There are different types and categories of cross-layering depending on the number of participating layers and the direction of the additional information flow. For instance, some cross-layer interactions commonly used are: the channel state information (CSI) in order to adapt throughput, the number of MAC layer retransmissions as a metric for the quality of the link, the quality of the incoming packet information as a metric for the routing algorithm, MAC layer error control as means of providing QoS at TCP or the priority of the message from Application layer on different schemes for better QoS. As it can be understood, the possibilities of optimisations and interactions are limitless and usually depend on the requirements set for each specific system. Our
Table 2.6: Cross-Layer routing protocols

<table>
<thead>
<tr>
<th>Routing Protocol</th>
<th>Cross-layer mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTR [52]</td>
<td>Link Residual Time calculated at PHY is exploited at NET</td>
</tr>
<tr>
<td>SBRS-OLSR [76]</td>
<td>SNR information from PHY is used to select MPR node at NET</td>
</tr>
<tr>
<td>MOPR [77]</td>
<td>Link stability from MAC information used in NET</td>
</tr>
<tr>
<td>R-AOMDV [78]</td>
<td>Transmission count from MAC with hop count at NET</td>
</tr>
<tr>
<td>DeReHQ [79]</td>
<td>Link reliability from MAC with End-to-End delay and hop count from MAC</td>
</tr>
<tr>
<td>PROMPT [55]</td>
<td>Two-way cross-layer: delay-aware MAC from NET information and relay selection using mobility information from NET</td>
</tr>
<tr>
<td>Barghi et al. [56]</td>
<td>Prediction of link stability at MAC using position/speed/direction from NET</td>
</tr>
<tr>
<td>Yang et al. [53]</td>
<td>Solve link asymmetries at MAC using hierarchical location service from NET</td>
</tr>
<tr>
<td>VTP [10]</td>
<td>NET information for bandwidth availability is exploited in Transport</td>
</tr>
<tr>
<td>Zhou et al. [57]</td>
<td>MAC persistence probability configured using Transport information and NET selects paths taking into account the rate</td>
</tr>
</tbody>
</table>

interest is mainly optimisations of the network layer, and more specifically use of cross-layer information from various layers to adapt routing decisions optimising a vehicular system performance. To this end, we present various proposals for cross-layering that could be used in a vehicular ad-hoc network (Table 2.6). A more extensive survey on cross-layer designs for VANETs can be obtained in [75].

Network with lower layers We start with cross-layer designs of NET and lower layers (MAC and PHY). The main objective of these approaches is to use channel quality information from PHY as means of link quality prediction based on which the routing protocol will perform the path selection. Protocols presented in 2.3.1.2 do not take into account the characteristics of the communication channel or the node’s utilization. They make use of simple metrics such as hop count, distance or enhancements of these, including navigation info. Using information about the received signal strength and arrival time of packets at the PHY, authors in [52] calculated the Link Residual Time (LRT) metric. This is an indicator of the remaining time that the specific link can be used for transmission. LRT is “exposed” to upper layers, such as routing. How-
ever, calculating LRT is not trivial. It requires removal of the noise from the data, estimation of the model parameters and finally renewing LRT. The advantage of this approach is that is generic; LRT can be used by any other upper layer. On the other hand, SBRS-OLSR \cite{76} is restricted to OLSR. Here, SNR information from PHY is used by the OLSR routing protocol in order to select the best MultiPoint Relay (MPR) nodes; the one with the highest SNR. These nodes are responsible for the topology broadcasting contrary to the initial OLSR where all nodes were broadcasting topology information. MOPR \cite{77} on the other hand uses movement information available at the MAC layer to predict the future positions of the relay nodes and calculate the “link stability” based on which the forwarding selection will be performed. Since this is MAC layer information, the upper network layer could be either a topological or a geographical protocol. It may seem similar to GPSR-L \cite{67} but in MOPR the position information is available at MAC whereas in GPSR-L it is directly available to NET thus it is not counted as cross-layer protocol. Another protocol that uses MAC information is R-AOMDV \cite{78}. It combines transmission count available at MAC and hop count available at NET to calculate its routing metric thus providing QoS based on the complete path and not only per link. A triple constrained routing protocol to provide better QoS in VANETs is DeReHQ (Delay-Reliability-Hop) \cite{79}. It is based on AODV but also considers the end-to-end Delay, link Reliability, and Hop count giving different priorities in these metrics. The previous routing protocols were based mainly on topological approaches, using hop count as their main route metric enhanced with some cross-layer information. However, as we mentioned in 2.3.1, geographic routing performs better in VANETs. PROMPT \cite{55} is a geographic routing protocol which has a bi-directional cross-layer design. It is developed for Vehicle-to-Infrastructure applications and provides (a) delay-aware routing through traffic statistics collected in MAC and (b) robust relay selection at MAC layer supported by mobility information from NET. Another geographic protocol is proposed in \cite{56}. It can predict the life-time of the communication link using stability metrics (positions, speed, direction) throughout the path thus selecting the more stable route to destination. Finally, a cross-layer design for heterogeneous MANETs is proposed in \cite{53} where nodes with different communication capabilities impose problems in routing due to link asymmetries. The solution
is given by the collaboration of MAC and NET using a hierarchical location service based on node density for the routing protocol and a multi-channel MAC to cope with link asymmetries. Such an approach could be useful for VANETs since there exist different types of nodes (vehicles, roadside units etc) which potentially have different capabilities.

**Network with upper layers** The second category that we present includes cross-layer protocols which use higher layer information to compute the path at the network layer. The objectives of these approaches are to provide different levels of service depending on the priority of the packet; e.g. safety applications require faster dissemination than infotainment. A novel cross-layer protocol for VANETs is the Vehicular Transport Protocol (VTP) [10]. It combines the transport layer with the network layer, using position-based routing to disseminate packets. Feedback information regarding bandwidth availability is passed from NET to transport layer using piggybacked ACK packets in order to provide congestion control. Another cross-layer design is proposed in [54] where the authors try to optimise TCP and GPSR [37] for vehicle mobility with adaptive interval of “HELLO” messages depending vehicle speed.

**Network with multiple layers** Finally, there are approaches that combine more than two layers. An example of these is presented in [57], where MAC, NET and Transport layer are jointly used for optimisation. With their joint algorithm they adapt persistence probability at the MAC layer using flow rate information. Then, using this information at the transport layer they adjust the source rate for rate control. At the end, routing is performed over the chosen link using the rate calculated before.

To summarise, the use of cross layer designs is the step forward in the protocol stack design from the strict layered approach. It is clear that the challenges imposed by the vehicular environment can not be solely faced by single-layered approaches. However, the amount of cross-layer information is an issue that should concern the researchers. The designs should be modular like in the OSI model but provide generic interfaces to other layers so that new protocols can be imported without a complete reconstruction of the protocol stack.
2.3.2 Location Services

Location Services (LS) can be implemented in a distributed manner through collaboration of network nodes or in a centralised manner like mobility management in cellular networks. Approaches like DREAM [80], LAR [81], and ETSI-LS [12] flood the entire network either with position updates or with queries, which causes severe overhead in the network. Alternatively, rendezvous based LS select a number of special nodes that serve as location service providers. In these schemes, the location updates and queries are not broadcast to every node, but they are directed to the location server nodes. A novel LS for VANETs is RLSMP [82]. It utilises mobility patterns to increase scalability and employs message aggregation for reduced overhead in querying. MG-LSM [83] also uses mobility information to group nodes travelling in the same direction and assigning one of them as the location server. This ensures a longer lasting association of a node with a single server; therefore, reducing the signalling overhead.

2.3.2.1 Taxonomy and characteristics

There are several architectures and approaches to categorise Location Services for VANETs. We follow the taxonomy presented in Figure 2.5. Most of the research is focused on the infrastructure-less LS where the mobile nodes play the role of location server. These LS are divided into two main categories. The first is the flood-based approaches where every node is a location server and either floods the whole network with position updates (e.g., DREAM [80]) or with position requests (e.g., LAR [81]). Such methods result in high volume of overhead and waste of resources which degrade the performance of the network. On the other hand, in rendezvous-based LS, some nodes play the role of the location server and hold position information for other nodes. This association is specified either by a hash function (in hash-based LS) or by groups (in quorum-based LS). In quorum-based LS, a node A sends location updates to a subset or region of the network, and the other nodes send requests for node A to a potentially different subset or region of the network. These two subsets are designed such that they intersect and the queries can be resolved. Such an example is the XYLS [84], where the updates are disseminated along the north-south direction and the request along the
Chapter 2. Background and Related Works

Figure 2.5: Location Service taxonomy

east-west. Hash-based LS use a strong hash function $H(x)$ (e.g., SHA-1, MD5) to map a node’s unique identifier (e.g., MAC address, IP address) to other nodes or regions that act as location servers for node $A$ [85]. This hash function is known to all network nodes, so when a node wants the position information of node $A$ it calculates $H(A)$ and sends requests to those nodes or regions. Node $A$ sends its updates to the same nodes respectively.

However, especially for VANET scenarios where infrastructure can be available, either with the use of cellular networks or designated Road Side Units (RSUs), a centralised architecture may be more suitable. Such approaches are generally used in cellular network to track the mobility of nodes over different base stations. An example is presented in [86], where RSUs are utilised to provide mobility management for nodes over GSM network. Position requirements for cellular networks are much lower than those needed for position-based routing in VANETs. In cellular networks, we only need the base station serving the node whereas in VANETs the exact location and more information (heading, velocity, etc.) are required.

Another characteristic that is important for location services is the locality of the servers [92, 93]. Some LS select the location servers randomly among all nodes, in which case some updates and requests might take a long time to reach the location...
server. To solve this problem other LS set constraints on the selection of the location servers so as to be near the serving node. Additionally, location servers can form a hierarchy that can help with the locality. The lowest level of the hierarchy reduces the cost of updates and local queries. If a query further away comes, it is resolved following the LS hierarchy. There are different update triggering mechanisms. There are LS that trigger the update periodically after a timer has expired, others with distance after the node has moved certain distance or crossed a boundary, and finally those which have both time and distance triggers.

### 2.3.2.2 Infrastructure-less LS

Considering the distinct characteristics of vehicular networks such as the lack of strict energy constraints and the high mobility of the nodes (vehicles) constrained by the road topology, several LS have been proposed. MALM [88] uses Kalman filters to calculate the current position of a node based on historical location information of other nodes. The approach is based on intelligent flooding of location information, which, however, results in high overhead. In [90], a Vehicle Location Service (VLS) is proposed that utilises digital map information to assign the location servers through a hash function. Another hash-based LS that uses ‘responsible sections’, such as traffic-light controlled intersections or bus stops, is presented in [89]. Vehicles travelling at those sections are assigned as location servers assuming that they slow down or stop for some time in these areas. These locations are known to all nodes a priori, so they can send their queries towards these locations by calculating a hash function to find the responsible ones.

A quorum-based LS (RLSMP) is presented in [91]. It divides the network in regions (called segments), which then are divided in cells. Nodes within a segment form a geographical cluster. Each cell has a leader (CL) that gathers all the location information for that cell. The nodes in the central cell of a segment play the role of location server for all the nodes of the cluster. They only get an aggregated summary of the location information stored in CL. However, this approach results in overhead in order to maintain location information in CL nodes that change dynamically. Also, even though the
clustering is static and based on the position of each node, determining the CL and transferring location information from the old CL to the new selected CL causes extra overhead. MG-LSM [83] is another quorum-based LS where the network is divided in regions each of which has a fixed location server called region head. Vehicles moving nearby and in the same direction form clusters and the cluster head takes the role of location server for the rest. To reduce the overhead, it reports to the region head only its exact location and for the rest of the nodes only membership information (unique identifiers of nodes in cluster) or changes of that. When a node wants the location of another node, it sends a query to its region head, which then searches for the cluster head associated with the node requested. The region head sends another query to the corresponding cluster head that replies with the exact location information of the node. This results to reduced overhead from update messages, but the query overhead and delay are increased.

2.3.2.3 Infrastructure-based LS

Two LS that utilise infrastructure are MRLSMP [87] and LEMM [86]. MRLSMP is a modified version of [91] to take into account the existence of infrastructure as location servers. The CL is designed to be a fixed RSU in order to reduce the overhead of transferring location information from an old CL to a new CL. However, the RSUs are not connected with any backbone wired network, and the LS is still a decentralised process. On the other hand, LEMM uses RSUs that are interconnected and there is only one centralised location server that can predict which RSU will serve each node. But, LEMM is not used as a mechanism to provide position information to unicast routing protocols running on vehicles but as a mobility management mechanism for cellular networks in highway scenarios.

2.3.3 Transport Protocols

In the introduction we presented the basic transport protocols for vehicular networks within the ITS protocol stack. Legacy protocols such as TCP and UDP can be used mainly for IP-based applications. BTP, a UDP-like transport protocol, is commonly
used with in the GeoNetworking. In addition, ITS specific transport protocols have been designed to cope with the characteristics of vehicular traffic, e.g., VTP [10] and VITP [11]. The proliferation of multiple network interfaces, e.g. wireless (IEEE 802.11p) and cellular (3GPP LTE), has spurred the development of multi-homing and multi-path technologies. Stream Control Transmission Protocol (SCTP) [94] is an IETF standard that provides multi-homing currently as backup when IP becomes unreachable. It supports a message oriented data delivery, full duplex, congestion control, with reliable and partial reliable (PR) data transfer. There are numerous extensions to this protocol that exploit the multiple network interfaces and paths available to increase throughput and reduce latency, e.g. concurrent multi-path transfer (CMT) approaches [95]. There are two approaches for scheduling in a multi-homed architecture as we review in sections 2.3.3.1 and 2.3.3.2 respectively. The first, is to use only one path to send packets and based on a utility function select the best path. The second, is to utilise the Concurrent Multi-path Transfer (CMT), where packets can be sent over multiple paths. These two approaches are examined in the following subsections.

2.3.3.1 Single Best Path Selection

In basic SCTP with multi-homing, only the primary interface is used. Alternate interfaces are considered secondary and used only in case of path failures; i.e. retransmit lost packets to increase probability of successful reception and transmission of new packets when primary is declared inactive (in this case the secondary is turned primary).

In [96], the selection of primary interface is based on two factors; round trip time (RTT) and estimated bottleneck bandwidth. RTT reflects the degree of congestion and the packet loss rate on a path. On the other hand, bandwidth provides information on which path can provide enough bandwidth for real-time communication. Similarly, Fracchia et al. [97] base their selection of the best path on the estimated bandwidth after a retransmission time-out occurrence. If a secondary path provides larger bandwidth than the current primary, the paths are swapped. This procedure is performed with a time hysteresis to avoid frequent path switches.

The problem with these proposals for single path, is that the use of greedy approach
on selecting the path with minimum RTT or maximum estimated bandwidth, makes users to select mainly one of the two networks, depending on the performance of the network at a single point of time and from the perspective of a single user. However, this causes inequalities in the individual cost, even though the total cost might be the same. In addition, the offloading ratio is not controlled and users may be forced to use a pricey network without a significant performance benefit.

2.3.3.2 CMT Best Path Selection

In CMT, by default, packets arriving from the upper layer are scheduled alternately on the available paths in a Round-Robin fashion, which is not efficient as reported in [98]. The throughput tends to get bounded by two times the throughput of the path with smaller bandwidth. This is as a consequence of two issues; (i) Packets must be scheduled on a path as long as the congestion window on the path is open, (ii) Sequence number holes in the receive window should be avoided. Since the receiver relays packets to the upper layer in sequence, existence of holes simply prevents the receive buffer from flushing. The simple Round-Robin scheduler does not address any of these two issues. The authors in [98] propose two strategies which try to schedule packets so to address the above two concerns.

- **Lazy**: Schedule packets on the current path until the congestion window exhausts and then switch to the other.

- **Smallest RTT**: Always schedule packets on the path with the smallest RTT that has congestion window open.

The Lazy scheduler tends to outperform the Round-Robin scheduler only when the disparity in the bandwidths of the two paths or the delays of the two paths is large. The Smallest RTT scheduler’s throughput is almost equal (closer than 90%) to the Ideal as it addresses the above two issues. In addition, this work considers the cost of using the network and has devised a utility function based on the throughput on each network (assuming WiFi and cellular). Their analysis showed that this function

\[ 	ext{Ideal} = \text{sum of the bandwidth of the two paths}. \]
is optimised when the maximum possible throughput of the WiFi and if that is no sufficient, send the remaining traffic on the cellular.

On the other hand, Cao et al. in [99] use a cross-layer QoS metric to select the path to send. The Cross-layer Model \((CM)\) metric combines the RTT of the path and the Frame Error Rate as follows.

\[
CM = \frac{1}{RTT \times \sqrt{FER}},
\]

where \(RTT = a \times \overline{RTT} + (1 - a) \times (T_{rx} - T_{tx} - \Delta T)\) denotes the round trip time on the path, \(\overline{RTT}\) is the current round trip time, \(T_{rx}\) the timestamp for receiving a packet, \(T_{tx}\) the timestamp for sending a packet, \(\Delta T\) the time interval for packet handling at receiver, \(a\) is a weighting factor usually 0.875 [100]. \(FER\) is the frame error rate at the MAC layer which can be calculated from the bit error rate \((BER)\)

\[
FER = 1 - (1 - BER)^n
\]

where \(n\) is the length of the packet in bytes. Fitzpatrick et al. in [101] use cross layer QoS metrics (SNR, RTT, packet loss and jitter) evaluated with the ITU E-model for path selection. However, this work can only be applied for VoIP traffic, since the E-model assesses user’s perception of speech transmission quality.

Xu et al. in [102] do not use the common RTT calculation as metric for path quality estimation because the acknowledgements may return from a different path. They also suggest that the calculation of RTT for every packet can not reflect the RTT variation process and estimate the trend of path quality. Hence, they propose to divide the total time of sending data into dissimilar periods in terms of the sending situation and calculate the quality metric for each of these periods as follows.

\[
Q = \frac{T_i - T_e}{\text{buffersize}},
\]

where \(T_e\) is the timestamp for the first packet send in that period, \(T_i\) is the timestamp for the last acknowledgement and buffer size is the allocated buffer size for that period. Using this metric and confidence intervals for selecting the sampling periods, they developed their proposed data distribution scheduler.
The optimal scheduling is modelled in [103] where Bui et al. model the problem of distributing data segments over multiple paths as a Markov chain, and formulate a Markov decision process to specify the scheduling policy. Following this, the On-line Policy Iteration (OPI) algorithm was proposed to approximate optimality. Although this work has substantial merit, it forgets two major constraints of the transport layer, that is, limited receive buffer and ordered data delivery. Without mention of these constraints, optimal throughput could be achieved, but unrealistically.

### 2.3.4 Propagation Models

In section 2.3.1, we discussed the importance of PHY and MAC layer information in order to predict the link quality for cross-layer schemes. Since research on ITS and vehicular communications is primarily based on simulation evaluations, due to the high cost of field experiments, proper care should be taken to model the channel characteristics. In VANETs, there are two types of propagation environments: highway and urban environments. In highway environment, nodes move mostly in straight lines, and usually a line-of-sight (LOS) model is appropriate. The challenge faced here is the high speeds of the vehicles that result in severe Doppler effect. In urban environments, on the other hand, the main challenge is the obstructing effect of buildings. Vehicles often do not have LOS with each other. An analytical model for the urban environments is presented in [104]. It takes into account both LOS and non-LOS (NLOS) components of the signal (Figure 2.6). Extensive field trials have been performed in the WINNER-II project, where a series of vehicular scenarios are defined and the appropriate channel models for link and system level simulations are investigated [105]. The difference between WINNER-II model for urban environments and TwoRay Ground propagation model [106], which does not consider fast-fading and obstacles, can be seen in Figure 2.7. In this figure, the dots represent vehicles and green lines the communication links. It is clear that the links are restricted to the road topology and close the corner of buildings.
2.3. Related Works

Figure 2.6: Propagation in urban scenarios with LOS and n-LOS components

Figure 2.7: Communication links in urban scenario simulated using NS-3 with and without considering the effect of building
2.3.5 Delay Modelling

Different models have been developed in the literature in order to evaluate a communication network in a systematic approach with respect to delay, predominantly using three methodologies: (a) Markov model, (b) Queueing Theory (QT), and (c) Network Calculus (NC). Markov models and QT can provide exact values for delay in a communication system, however may struggle to provide results for complex systems. NC provide an alternative to the classical queueing theory for analysing backlog and delay in communication networks. It uses more relaxed characterisation of distributions, which are defined by violation probabilities of arrival and service processes. Thus, providing bounds for the delay and backlog compared to the exact analysis of queueing theory, which may not be tractable for complex real systems. Therefore, as explained in next paragraphs, NC models divide complex systems into smaller ones analysed with QT or Markov models, which are then extended using NC theorems for the complete system.

For short range communications, most of the existing studies using Markov models are based on extensions of [107] for saturated data traffic case in IEEE 802.11-based, single hop scenarios. We do not review those, as they do not apply to multi-hop and non-saturated scenarios we investigate in this work. For non-saturated scenarios, there is a number of related works in the literature as follows. In [108], Felemban et al. have introduced a tight and accurate model for IEEE 802.11 DCF. The model accounts for channel state during the backoff countdown process, which increases the collision probability estimation accuracy. Using an iterative algorithm to compute the binomial distribution for the contenting nodes, the authors extend the saturated model to unsaturated cases. However, this work considers only single-hop scenarios. On the other hand, [109] introduces a model where a wireless node is represented by a discrete time $G/G/1$ queue. The service time distribution for the queues is derived by accounting for a number of factors including the channel access delay due to the shared medium, impact of packet collisions, the resulting backoffs as well as the packet size distribution. This is extended for arbitrary packet size distributions and queue priorities as in IEEE 802.11e standard. Still this model only considers single-hop scenarios. One of the
most recently developed models for IEEE 802.11 DCF is introduced in [110], which combines the Markov modelling with QT. The backoff transitions are considered a Markov renewal process and the service is characterised by a $M/G/1$ queue. The Markov renewal process model simplifies the derivation of the closed form solution for the probability that each station attempts to transmit in a slot. Furthermore, two recent studies model IEEE 802.11 using NC methodology [111, 112]. The first model does not provide an analytical form for the calculations of the upper bound of the service curve, which is evaluated numerically according to a heuristic algorithm. Moreover, both of them are restricted to saturated single hop scenario deriving their service curves from Kumar et al. IEEE 802.11 model [113]. On the other hand, Gupta et al. [114] present a model for lower bound on delay in multi-hop scenarios. The aim in that work is to develop a delay-efficient scheduler. It is claimed that lower bound technique captures the effect of interference and statistical multiplexing of packets in the system. Jiao et al. [115] use basic probability theory and NC to analyze the delay a packet experiences at each hop along a path. Then, end-to-end delay is calculated through summing up the per-hop delay along the path. However, it has been shown in [116] that the complexity of a system is proportional to $O(n^2 \log n)$ when analysed hop-by-hop, compared to $O(n \log n)$ when analysed as one system according to the theorem presented later in the thesis.

Mathematical models of end-to-end delay in cellular networks have not been adequately investigated. A model for 3G cellular technology in [117] analyses the delays contributed from RLC and PHY layers on IP packets based on stochastic models. A semi-analytical Markov model of MAC layer of LTE is presented in [118], where the average delay of packets can be derived but it is limited to uplink. Gao et al. [119] and Zhang et al. [120] have used NC methodology to calculate the delay bounds in the LTE network. The work described in [119] is restricted to the air interface model of LTE and a specific case of applications related to the Internet of Things. It models the LTE service with a simple Gilbert-Elliot channel without considering delays in the evolved packet core (EPC) and assumes constant traffic from a sensor node to a remote-host. On the other hand, [120] presents a more generic LTE architecture, which considers a MIMO air interface as well as an EPC with multiple routers and strict priority scheduling. Each
component is modelled with a stochastic service and they are combined in a single system. Therefore its complexity is $O(n \log n)$ as explained in [116]. The arrival traffic consists of both real-time and non-real-time flows.

2.4 Summary

In this chapter, the network architecture of Intelligent Transport Systems was presented and related work was reviewed. Several applications for ITS have been realised and many more will be developed in the future, each one with different characteristics and communication requirements. Numerous communication technologies are promoted in vehicular networks, that fulfil specific application requirements.

A vast range of routing protocols are proposed for VANETs, some of which are reviewed in this thesis. The challenges opposed by the characteristics of VANETs favour the use of geographical routing against topological, hierarchical or flooding. However, using position information for the forwarding is not enough. It has to be enhanced with navigation information since the nodes are vehicles and their mobility is constraint by the road network. Geographical routing comes with a weakness; the local maximum problem. An appropriate recovery strategy should be employed to cope with this and since the nodes move with relative high velocities rapidly changing the network topology, the carry-n-forward mechanism is the most suitable. In Intelligent Transportation Systems, there are different applications that require reliable wireless communications with certain QoS constraints. Simple geographical routing fails to meet these requirements, therefore cross-layer designs have been proposed. In order to simulate realistic VANET scenarios and assess the impact of cross-layering, proper propagation models have to be modelled. In Table 2.7 a summary of all protocols studied in this thesis are presented showing their type (topological, geographical etc), the metric that is employed by the protocol for the path selection, whether or not a navigation system is used, the recovery mechanism employed to cope with the local maximum problem and finally if it is a cross layer design, which layers are coupled.

Moreover, since geographical routing requires a location service, the design of a suitable one is vital. The review of related location services for VANETs has shown that
infrastructure-based LS are good candidates, since they can exploit infrastructure already existing in urban areas and mix different access technologies in order to increase the performance of the network.

The use of multiple network interfaces, i.e. multi-homing, can increase the reliability of the network by providing backup paths for communication, or the throughput by utilising multiple paths simultaneously. The deployment of a transport protocol, which can operate and take advantage of multi-homing provides a significant advantage in a communication system. SCTP is an IETF protocol that supports multi-homing and most of the reviewed proposals are based on it. However, these do not consider the cost of using each network or the fairness among individual users.

The analytical evaluation of a communication system can provide information for the limits of an architecture that can not be easily simulated or tested in real systems. Delay is a key metric for the performance of a system, and a number of models have been proposed to mathematically trace it. However, models for end-to-end delays in ad-hoc networks are limited to single hop scenarios, and in cellular are inadequately investigated.
### Table 2.7: Summary of routing protocols for VANETs

|------------------|-------------------------|-----------------------------------------|------------|-----------------|-------------|
| OLSR [28], DSDV [29] | topological  
(proactive)                  | hop count                              | x          | x               | x           |
| AODV [30], DSR [31] | topological  
(reactive)                  | hop count                              | x          | x               | x           |
| TORA [32], ZPR [33] | topological  
(hybrid)                  | hop count                              | x          | x               | x           |
| HRS [34] | topological  
(hierarchical)                  | hop count                              | x          | perimeter routing | x           |
| GPSR [37] | geographical                  | distance (euclidean)                  | x          | x               | x           |
| Finn et al. [40] | geographical                  | distance (euclidean)                  | x          | x               | x           |
| MFR [39] | geographical                  | distance (most forward radius)         | x          | x               | x           |
| NFP [42] | geographical                  | distance (nearest forwarding progress)  | x          | x               | x           |
| Nelson et al. [43] | geographical                  | distance (random positive progress)    | x          | x               | x           |
| Compass [44] | geographical                  | angle                                  | x          | x               | x           |
| CAR [58], GPSRJ+ [59] | geographical                  | distance plus number of intersections  | ✓          | x               | x           |
| GPCR [60] | geographical                  | distance plus number of intersections  | ✓          | right hand rule | x           |
| GyTAR [38] | geographical                  | distance plus traffic density and intersections | ✓          | Carry-n-forward | x           |

Continued on Next Page ...
Table 2.7 – Continued

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>VADD [61]</td>
<td>geographical</td>
<td>distance plus prediction, road speed limit</td>
<td>✓</td>
<td>Carry-n-forward</td>
<td>x</td>
</tr>
<tr>
<td>A-STAR [62]</td>
<td>geographical</td>
<td>distance plus prediction</td>
<td>✓</td>
<td>Re-route using different anchor points</td>
<td>x</td>
</tr>
<tr>
<td>AGF-GPSR [63]</td>
<td>geographical</td>
<td>distance plus number of intersections</td>
<td>✓</td>
<td>Re-route using different anchor points</td>
<td>x</td>
</tr>
<tr>
<td>Optimized GPSR [64]</td>
<td>geographical</td>
<td>distance plus number of intersections</td>
<td>✓</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>MP2R [65], MAGF</td>
<td>geographical</td>
<td>distance plus prediction</td>
<td>✓</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>GPSR-L [67]</td>
<td>geographical</td>
<td>distance plus lifetime</td>
<td>✓</td>
<td>perimeter routing random delay on packet retransmission</td>
<td>x</td>
</tr>
<tr>
<td>CGGC [46]</td>
<td>geographical</td>
<td>distance (euclidean)</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>MoVe [47], SAR [48]</td>
<td>geographical</td>
<td>distance (euclidean)</td>
<td>x</td>
<td>Carry-n-forward</td>
<td>x</td>
</tr>
<tr>
<td>Gasari et al. [49], Basagni et al. [50]</td>
<td>geographical</td>
<td>distance plus colour</td>
<td>x</td>
<td>Colouring</td>
<td>x</td>
</tr>
<tr>
<td>Improved GPSR [51]</td>
<td>geographical</td>
<td>distance (euclidean)</td>
<td>x</td>
<td>Re-route using different anchor points</td>
<td>x</td>
</tr>
<tr>
<td>LTR [52]</td>
<td>geographical</td>
<td>distance plus LTR</td>
<td>x</td>
<td></td>
<td>NET + PHY</td>
</tr>
</tbody>
</table>

Continued on Next Page . . .
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SBRS-OLSR [76]</td>
<td>topological</td>
<td>hop count</td>
<td>×</td>
<td>×</td>
<td>NET + PHY</td>
</tr>
<tr>
<td></td>
<td>(proactive)</td>
<td>hop count plus link stability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOPR [77]</td>
<td>topological</td>
<td>hop count plus Tx count</td>
<td>×</td>
<td>×</td>
<td>NET + MAC</td>
</tr>
<tr>
<td>R-AOMDV [78]</td>
<td>topological</td>
<td>prioritized delay, reliability and hop</td>
<td>×</td>
<td>×</td>
<td>NET + MAC</td>
</tr>
<tr>
<td></td>
<td>(reactive)</td>
<td>count</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DeReHQ [79]</td>
<td>topological</td>
<td>distance plus MAC statistics</td>
<td>×</td>
<td>×</td>
<td>NET + MAC</td>
</tr>
<tr>
<td></td>
<td>(reactive)</td>
<td>distance plus mobility metrics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROMPT [55]</td>
<td>geographical</td>
<td>distance (euclidean)</td>
<td>×</td>
<td>×</td>
<td>NET + MAC</td>
</tr>
<tr>
<td>Barghi et al. [56]</td>
<td>geographical</td>
<td>distance plus bandwidth availability</td>
<td>×</td>
<td>×</td>
<td>NET + MAC</td>
</tr>
<tr>
<td>Yang et al. [53]</td>
<td>geographical</td>
<td>distance (euclidean)</td>
<td>×</td>
<td>×</td>
<td>NET + MAC</td>
</tr>
<tr>
<td>VTP [10]</td>
<td>geographical</td>
<td>distance plus rate and MAC info</td>
<td>×</td>
<td>×</td>
<td>NET + Transport</td>
</tr>
<tr>
<td>Chen et al. [54]</td>
<td>geographical</td>
<td></td>
<td>×</td>
<td>×</td>
<td>NET + Transport</td>
</tr>
<tr>
<td>Zhou et al. [57]</td>
<td>geographical</td>
<td></td>
<td>×</td>
<td>×</td>
<td>NET + MAC + Transport</td>
</tr>
</tbody>
</table>
Chapter 3

Cross-Layer Optimised Geo-Routing

3.1 Introduction

This chapter presents the proposed geo-routing protocol named Cross-Layer, Weighted, Position-based Routing (CLWPR). In Subsection 2.3.1 we reviewed existing proposals for routing in vehicular networks. This review has shown that position-based routing protocols, which consider navigation outperform topology-based and position-based protocols without navigation. In addition, cross-layer information can significantly enhance the next-hop selection process. Therefore, our proposal is a cross-layer position-based protocol with navigation information. To the best of our knowledge, none of the protocols we reviewed considers the related factors e.g., mobility information, within a coherent framework, nor investigate the effects of communication and environment parameters on forwarding decisions. To this end, we propose CLWPR to be used in vehicular environments. With the use of cross-layer information from physical (PHY) and data link (MAC) layers, the proposed algorithm is able to estimate the link quality, that is taken into account in the routing decision. Moreover, information about node’s position, speed and heading is used by a prediction scheme in order to have more accurate position information. In addition, an adaptive HELLO message exchange mechanism among neighbour nodes is used to reduce the signalling overhead required for routing.
process. Navigation information, regarding the roads that vehicles are travelling, and heading are also considered in the forwarding selection in order to reduce end-to-end delay. To cope with frequent link failures and network segmentations, mainly in sparse networks, a carry-n-forward mechanism is employed. To investigate the effects of the aforementioned parameters, an Analytic Hierarchy Process (AHP) [121,122] is adopted in our proposed protocol for making routing decisions.

The remainder of this chapter is comprised of six sections: (3.2) description of the proposed routing protocol design; (3.3) an example of CLWPR operation; (3.4) discussion of the routing metric that is used for forwarding function; (3.5) specification of the AHP approach used in this thesis; (3.6) evaluation of the proposed protocol; and (3.7) discussion on the findings of the simulation campaign and other issues.

### 3.2 Protocol Design

CLWPR is a distributed unicast, multi-hop, cross-layer protocol based on opportunistic forwarding. Unlike reactive routing protocols, it does not rely on route discovery. The selection of the next hop, during the forwarding process, is performed based on calculation of a decision metric for all neighbour nodes, called weight in this thesis. The forwarding algorithm at the heart of the proposed routing protocol can be visualised with the flowchart in Figure 3.1. The algorithm first checks if the destination’s position information is known. If not, a request is sent to a location service. When the information becomes available, the node calculates the weight of all its neighbours based on local information (neighbour list), as it will be described in Subsection 3.4. If the forwarding node faces the local maximum problem, namely it has the least weight among its neighbours, the packet is stored locally until a neighbour with less weight is found or until a timer expires. The forwarding algorithm relies on the neighbour discovery mechanism, which is based on 1-hop “HELLO” messages that every node periodically broadcasts. As summarised in Table 3.1 these messages include positioning information (position, velocity) of the broadcasting node, the node’s MAC related information, and the number of cached packets due to local maximum. Each node updates its local list of neighbours with the information learnt from these messages. In
addition, upon receipt of a “HELLO” message, a node calculates the SINR value of the received message and stores it with the rest neighbouring information. Then, it counts the consecutive “HELLO” messages received from the same neighbour, as an indicator of neighbour reliability. In order to reduce the overhead of these broadcast messages, CLWPR employs a dynamic broadcasting scheme where the inter-arrival of the packets varies according to vehicle’s speed. The information used for forwarding purposes comprises of three basic components: mobility, link quality and node utilisation.

Table 3.1: HELLO Message Information

<table>
<thead>
<tr>
<th>Information Carried</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node Position</td>
<td>The position co-ordinates (x,y)</td>
</tr>
<tr>
<td>Node Velocity</td>
<td>The velocity co-ordinates (x,y)</td>
</tr>
<tr>
<td>MAC Frame Error rate</td>
<td>The average number of collisions in unit of time</td>
</tr>
<tr>
<td>C’n’F indicator</td>
<td>The number of cached packets due to local maximum</td>
</tr>
</tbody>
</table>

Figure 3.1: CLWPR forwarding algorithm
3.2.1 Mobility related information

Unlike greedy forwarding policies and other position based protocols \cite{12,37,61,63}, CLWPR does not calculate the minimum geographic distance between two nodes. Instead, it determines the actual distance that a vehicle would have to travel in order to reach the destination, called *curvemetric* distance in this thesis. The motivation for this design decision originates from the fact that the nodes are vehicles and as such their movement is restricted within the read boundaries. Packets have to be forwarded alongside roads to avoid propagation obstacles, such as buildings, that might block the direct path among communicating nodes. Thus, the distance of two vehicles is better described by the distance based on the road network layout rather than their minimum geographical (Euclidean) distance. In contrast to other protocols in \cite{38,58,60,69}, CLWPR does not use “anchor” nodes at intersections like protocols through which a packet has to be forwarded, thus it reduces the overhead for identifying these nodes. In order to be able to calculate the *curvemetric* distance, electronic maps (e-maps) should be available from the vehicles; e.g., navigation systems. Then the road that a vehicle is travelling on can also be identified. As discussed, when a message forwarded along the road that the destination is travelling we maximize the probability to have LOS communication. Such a selection is performed close to junctions where more vehicles can be accessed. Then those vehicles travelling along destination’s road and approaching it are preferred. More frequent “HELLO” messages can provide more accurate and up-to-date information of a node’s position. However, such an approach increases network overhead. In our protocol, we use the information gathered from “HELLO” messages, such as position, speed and heading (extrapolated from the velocity vector), to predict future positions of a node in order to reduce the frequency of broadcasted “HELLO” messages. In addition a dynamic broadcasting scheme is proposed where the interval varies with vehicle’s velocity to further reduce overhead.

3.2.1.1 Dynamic HELLO Broadcast scheme

The frequency of HELLO messages plays a significant role in the performance of geographic routing protocols. Most of them have fixed, very frequent inter-arrival which
causes increased overhead. Less frequent broadcasts in high vehicles’ speeds decrease the performance. We investigate the impact of node’s velocity and HELLO interval on the performance of the routing protocol, in terms of packet delivery ratio with and without position prediction (please refer to Section 3.6 for scenario details). It can be seen from Figure 3.2 that more frequent HELLO broadcasts, increase PDR especially without the use of prediction. However, such increase of HELLO frequency means also increase in network overhead caused from HELLO packets.

We notice the correlation between vehicle speed and inter-arrival rate and the correlation between average vehicle speed and vehicle traffic density (in sparse traffic vehicles tend to travel faster than in congested areas). Therefore, we propose a dynamic broadcast scheme, where the inter-arrival varies between a maximum rate when vehicles travel faster than a certain maximum speed and a minimum rate for nodes traveling slower than a certain minimum speed. One way of doing this is with the use of step
function as follows:

\[
\text{HELLO Interval} = I = \begin{cases} 
I_{\text{max}}, & \text{if } V \leq V_{\text{min}} \\
I_{\text{max}} - i \cdot Thr, & \text{if } V_{\text{min}} + i \text{ step} < V \leq V_{\text{min}} + (i + 1)\text{step} \\
I_{\text{min}}, & \text{if } V > V_{\text{max}}
\end{cases}
\]  

(3.1)

where Threshold \((Thr)\) can be calculated depending on the number of steps \(i\) (granularity) we want to use. The minimum and maximum interval, \(I_{\text{min}}\) and \(I_{\text{max}}\) respectively, with the speed range \((V_{\text{min}}, V_{\text{max}})\) are a matter of design decision. For example, the interval can be related to the ETSI’s Cooperative Awareness Message (CAM) sending rate with minimum 0.1 second and maximum 1 second \([123]\). The speed range could be then extrapolated from an analysis similar to Figure 3.2, where the maximum performance would be selected for particular \{speed, interval\} pair.

3.2.2 Link quality related information

Since the communication links in vehicular environments are highly variable and perhaps short lived due to the dynamic nature of the network, cross-layer information from PHY and MAC layers will help select more reliable forwarding nodes. However, the cross-layer approaches discussed in Section 2.3.1.3 are not suitable for all VANETs scenarios since they rely on existence of infrastructures (e.g., \([55]\)) or require complex calculations (e.g., \([52]\)). Therefore, assuming channel reciprocity, we propose to use the SINR value of the received “HELLO” messages as a metric of link quality. Moreover, we use the MAC layer errors, e.g., contention errors, as another metric that will contribute to further increase the reliability of our routing protocol. The number of consecutive received “HELLO” messages is used as an indicator of neighbour node reliability.

3.2.2.1 CSI Function

The selection of a channel quality function is based on the characteristics of message dissemination. Interference is relatively high in VANETs and nodes that are located close to the border of the communication range experience more adverse phenomena than those near the centre. Thus, SINR value of the received messages at the border
is lower. Our approach is to select nodes far enough from the source, but within good communication range of the source node. This is achieved by selection of the appropriate SINR threshold ($SINR_{th}$) for which the weight is minimised (Figure 3.3). Nodes with lower SINR than $SINR_{th}$ will have higher weight because they are closer to the border and the probability that the message will be dropped is increased. Also, nodes with higher SINR value will have higher weights to give them higher forwarding priorities. We have selected the following CSI function that fulfils our requirements.

$$CSI = \begin{cases} 
  ax^2, & \text{if } SINR_r \leq SINR_{th} \\
  b/x, & \text{if } SNIR_r > SINR_{th}
\end{cases}, \quad (3.2)$$

where $a/b = 1/x^3 \big|_{x=(SINR_{th}-SINR_{min})}$ and $x$ is the difference between the obtained $SINR_r$ value and the lowest SINR at the border of the communication ($SINR_{min}$).
3.2.3 Node utilisation related information

As mentioned in Section 2.3.1, geographical routing protocols suffer from the local maximum problem, especially in low density networks. This is also the case with CLWPR. We address this problem by adopting a carry-n-forward mechanism. This selection is based on the fact that, in VANETs, neighbour nodes vary frequently due to the high and constraint mobility. Such updates may result in new nodes that solve the local maximum problem. Therefore, it is preferred to cache the packet shortly than start a recovery mechanism like perimeter routing that would forward the packet away from the destination or drop the packet. This will occasionally result in higher end-to-end delays. However, in order to reduce the effect of caching packets locally, we “penalise” those nodes with extra weight related to the number of cached packets.
3.3 CLWPR Operation With an Example

An example demonstrating CLWPR forwarding algorithm can be viewed in Figure 3.4. In this example there is one source (S) and two destination (D1 and D2) nodes. We assume that every node knows the position of the destinations through a Location Service mechanism and its neighbours through the “HELLO” message exchange mechanism as listed in Table 3.2. When S wants to send a packet to either destinations it looks into its neighbour list. If the Euclidean distance was used, S would be selected for D2 and B selected for D1. In both cases, such selection would not be efficient due to the local-maximum problem already discussed. With the use of curvemetric distance, however, node A will be selected without having to identify intersections and anchor points. The next hop selection from node A towards both destinations will be one of nodes C and D, which both have the same curvemetric distance. Since, node C moves towards the destinations, whereas D is travelling away, it is preferred as the second hop. Then, node C will have to select the next-hop for the destinations among nodes E – I. Node I is the closest node to D1, but at the edge of the communication range of C, and therefore it has a high probability of dropping a packet. In addition, this high traffic intersection will cause high contention among the nodes, so proper caution should be taken in the next-hop selection. For D1, the nodes with the least weight will be selected and the packet will be delivered within 4 hops. However, D2 is out of the communication range of all nodes, which again results in local-maximum problem. Therefore, carry-n-forward.
mechanism is employed. If a different recovery mechanism was used (e.g. perimeter
routing), packets would be forwarded away from the destination. If, for example, node
G is selected as next-hop for D2, the packets will need to be cached. To avoid losing
packets due to increased data-flow to D2 and buffer overflow on node G, a different
node will be selected when the number of cached packets is significantly increased.

3.4 CLWPR Weighting Function

A forwarding node $i$ computes the weight of neighbor node $j$ with respect to routing
to destination node $k$, denoted by $W^{(k)}_{i,j}$, as follows:

$$W^{(k)}_{i,j} = f_M M^{(k)}_{i,j} + f_L L_{i,j} + f_I \Gamma_j,$$  \hspace{1cm} (3.3)

where:

- $f_X$ indicates the relative importance of parameter $X$ in making forwarding de-
cisions. These factors are calculated using the AHP algorithm as presented in
Subsection 3.5.
- $M^{(k)}_{i,j}$ accounts for the impacts of mobility on routing decisions, given by:

$$M^{(k)}_{i,j} = f_D D^{(k)}_{i,j} + f_R R_{j,k} + f_P P_{j,k},$$  \hspace{1cm} (3.4)

where $D^{(k)}_{i,j}$ is the normalised curvemetric distance of neighbor $j$ from destination
node $k$ calculated at node $i$ as follows.

$$D^{(k)}_{i,j} = \frac{D_{j,k} - D_{i,k}}{r},$$  \hspace{1cm} (3.5)

Here $D_{i,k}$ and $D_{j,k}$ are the curvemetric distance of forwarding node $i$ and neigh-
bour $j$ from the destination $k$ based on the current position of the corresponding
nodes, and $r$ is the nominal communication range of a node. The current posi-
tion is estimated using the knowledge acquired from the most recent “HELLO”
message and the assumption that the node does not change direction between
two consecutive “HELLO” messages. If neighbour node $j$ and destination node $k$ are on the same road $R_{j,k} = 0$, and $R_{j,k} = 1$, otherwise. Further $P_{j,k}$ indicates whether node $j$ will be in a closer position or further position to node $k$ based on their current travelling paths. Assuming that the destination node $k$ is fixed for a certain period of time, $P_{j,k}$ can be quantified by the cosine of the angle $\theta$ between the velocity vector of node $j$ ($\vec{V}_j$) and the vector starting at node $j$ towards node $k$ ($\vec{JK}$) as follows:

$$P_{j,k} = -\cos(\theta) = -\frac{\vec{V}_j \cdot \vec{JK}}{||\vec{V}_j|| \cdot ||\vec{JK}||},$$

(3.6)

- $L_{i,j}$ represents the link information between forwarding node $i$ and neighbour node $j$, given by:

$$L_{i,j} = f_C CSI_{i,j} + f_Ma M_j + f_NR_{i,j},$$

(3.7)

where $CSI_{i,j}$ represents the quality of the channel between forwarding node $i$ and neighbour node $j$. $M_j$ indicates the level of contention in the area close to the neighbour node $j$ represented by the average number of collisions, and $NR_{i,j}$ represents the reliability of the neighbour node $j$. This is calculated based on the number of consecutive “HELLO” messages that node $i$ received from a neighbour $j$ on expected intervals, denoted by $H_c$. We choose the following values for $NR_{i,j}$ between 0, which indicates a highly reliable node, and 1, which indicates a less reliable node:

$$NR_{i,j} = \begin{cases} 
1 & \text{, if } H_c \leq 2 \\
0.5 & \text{, if } 2 < H_c \leq 4 \\
0 & \text{, if } H_c > 4 
\end{cases}$$

(3.8)

- $\Gamma_j$ is the ratio of number of carry-n-forward packets at node $j$ to the queue size.

We take into account this parameter to reduce the chance of selecting next hop nodes that are facing local maximum problem.

Determining a good set of $f_X$ parameters, which will optimise the performance of CLWPR, or providing insights of the effects of aforementioned parameters on the net-
work performance, is a non-trivial problem. In this direction, an AHP based methodology is described in the next subsection, to systemically approach this problem in a typical urban environment.

### 3.5 Analytic Hierarchy Process (AHP)

AHP \[121, 122\] is a general approach that has been used in multi-criteria decision analysis, similarly to our approach in Subsection 3.4. The AHP decomposes the decision problem into elements, according to their common characteristics, and hierarchy levels. The top level consists the “goal” of the problem and the rest levels correspond to relevant criteria and sub-criteria. AHP is then used to evaluate the relative importance between the criteria. Our AHP-based approach is implemented in three steps, following the methodology in \[121\].

#### 3.5.1 Description of problem as a hierarchy

We describe the multi-criteria forwarding decision, defined in \[3.3\], with an AHP hierarchy as shown in Figure 3.5. The goal of our approach is to calculate the weight of all individual nodes from the neighbour list, and subsequently to select the neighbour with the minimum weight. The first level of hierarchy includes the high level decision criteria: mobility, link quality, and node utilisation. The second level further expands these criteria into more detailed sub-criteria corresponding to \[3.4\] and \[3.7\].
Table 3.3: Scales of Pairwise Comparison

<table>
<thead>
<tr>
<th>Importance</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Equally Important</td>
</tr>
<tr>
<td>3</td>
<td>Moderate Importance</td>
</tr>
<tr>
<td>5</td>
<td>Strong Importance</td>
</tr>
<tr>
<td>7</td>
<td>Extreme Importance</td>
</tr>
<tr>
<td>9</td>
<td>Extremely More Important</td>
</tr>
<tr>
<td>2,4,6,8</td>
<td>Intermediate values between adjacent scales</td>
</tr>
</tbody>
</table>

3.5.1.1 Construction of pair-wise comparison matrix

The next step is to construct a comparison matrix for each level, denoted by \( C \).

\[
C = \begin{bmatrix}
    c_{1,1} & c_{1,2} & \cdots & c_{1,n} \\
    c_{2,1} & \ddots & \vdots & \vdots \\
    \vdots & \ddots & \ddots & \vdots \\
    c_{n,1} & \cdots & \cdots & c_{n,n}
\end{bmatrix},
\]

where \( n \) is the total number of criteria as each level, and \( c_{i,j} \) represents the relative importance of criteria \( i \) to \( j \) constrained by the following rules: \( c_{i,j} > 0; c_{i,j} = 1/c_{j,i}; c_{i,i} = 1 \) for \( \forall i \). The exact values of \( c_{i,j} \) will be assigned according to the convention in Table 3.3 [121]. Note that each row and column of this matrix corresponds to one of the decision criteria given in (3.3), (3.4), and (3.7) in the specific AHP decomposition.

3.5.2 Calculation of \( f_X \) parameters

According to the AHP approach, we first need to normalize the comparison matrix, \( C \), as follows:

\[
c_{i,j} = \frac{c_{i,j}}{\sum_{i=1}^{n} c_{i,j}}.
\]

Then, if the \( f_X \) parameter, related to criteria \( X \), corresponds to row and column \( k \) in our comparison matrix, it can be computed as follows:

\[
f_X = \frac{\sum_{j=1}^{n} c_{k,j}}{n}.
\]
3.6 Performance Evaluation of CLWPR

In this section, we present the performance evaluation of the proposed protocol. There are several metrics that can be used to measure the performance of a routing protocol, but the most widely accepted ones are Packet Delivery Ratio (PDR), End-to-End Delay (E2ED), and overhead introduced by the routing protocol. PDR is calculated as the ratio of delivered data packets against the total number of sent data packets, E2ED is the average end-to-end delay experienced by received data packets, and overhead is specified as the ratio of the total amount of information used for signalling against the total size the information exchanged in the network.

Our simulation model examines two scenarios. In Scenario 1, we consider an urban area consisting of a Manhattan grid road network with 16 intersections, based on the reference area depicted in Figure 3.4 with edge size 2000m. This is a well-known benchmark simulation scenario type used in literature [124]. The mobility traces of this scenario are generated using Bonnmotion tool [125] for different vehicles densities within the speed limits for urban areas (avg speed 50km/h, std.dev. 5km/h). Scenario 2 simulates a real city environment with traces obtained from [63] for the urban area of “Unterstrass” in Zurich (Figure 3.6). The propagation model in [105] is considered in this thesis, which is suitable for both LOS and non-LOS communications. We consider two types of communications, namely vehicle-to-infrastructure (V2I) and vehicle-to-vehicle (V2V). For the first type, we employ a static RSU at the centre of the reference area to act as a server that consumes all the data traffic. For the second type, we have simultaneous connections between randomly selected moving vehicles. The vehicles and RSU are equipped with IEEE 802.11p communication units. The nominal communication range of the nodes is 500m when there is no obstacle in the LOS communications path. The simulation platform is NS-3, where the CLWPR model was developed. The outcomes of the simulations are averaged over a set of independent runs to produce the graphs as mobility and random backoff processes impact the results. The most important simulation parameters are summarised in Table 3.4.

In the remainder of this section, first we evaluate the impacts of different parameters in (3.3), (3.4), and (3.7) on the performance of the proposed routing protocol.
3.6. Performance Evaluation of CLWPR

Table 3.4: Simulation Parameters for CLWPR evaluation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal Comm. Range</td>
<td>500m (at line-of-sight) [105]</td>
</tr>
<tr>
<td>Number/Type of Connections</td>
<td>15 UDP, V2V &amp; V2I</td>
</tr>
<tr>
<td>MAC/PHY protocol</td>
<td>IEEE 802.11p, 6Mbps</td>
</tr>
<tr>
<td>Routing protocols</td>
<td>CLWPR, ETSI-GF, AGF, GyTAR</td>
</tr>
<tr>
<td>HELLO interval</td>
<td>1sec (*)</td>
</tr>
<tr>
<td>Caching Limit</td>
<td>5sec</td>
</tr>
</tbody>
</table>

(*) for CLWPR it is dynamic from 1-3sec

using AHP (Section 3.6.1). This aims to find the optimal $f_X$ values for CLWPR. Then, in Section 3.6.2, we compare the performance of the optimized CLWPR protocol with the ETSI proposal (ETSI-GF) that relies only of Euclidean distance between nodes, an advanced implementation of greedy forwarding, namely AGF, that supports carry-n-forward mechanism and a prediction policy (similar to VADD), and GyTAR. Appendix A provides supplementary results for an initial evaluation of CLWPR.

3.6.1 Impact of Forwarding Parameters

The use of AHP and particularly by means of the hierarchy structure, can group similar parameters, e.g. distance, proximity, road, and examine the effect of them as a whole
Chapter 3. Cross-Layer Optimised Geo-Routing

(mobility group) against other groups and individually among the same group. The output we get is not just a set of optimal parameters for our protocol, but insights of how such parameters affect the forwarding process. We carry out a number of simulations in order to find a set of appropriate $f_X$ parameters in (3.3), (3.4), and (3.7) that will optimise the performance of CLWPR for a typical urban area, represented by Scenario 1. This type of scenario with frequent LOS/non-LOS transitions poses one of the most adverse environments for a VANET routing protocol. We have performed a comprehensive analysis using an extensive number of configurations in order to suggest the most effective set of $f_X$ parameters. However, we only present the results for a typical medium density ($\sim 10$ veh/km) scenario with an average node speed of 50km/h and standard deviation 5km/h, assuming V2V communication traffic. The observed trends for other configurations not presented here, were very similar to those presented in this subsection.

First, we evaluate the impacts of $f_M$, $f_L$, and $f_T$ in (3.3). This indicates of the importance of mobility, utilisation and link quality information in the forwarding decision. Following the AHP methodology, we construct the comparison matrix for these parameters as follows:

$$
\begin{bmatrix}
1 & c_{M,L} & c_{M,T} \\
c_{L,M} & 1 & c_{L,T} \\
c_{T,M} & c_{T,L} & 1
\end{bmatrix},
$$

(3.12)

where $c_{L,M}$ represents the relative importance of Link quality related information to Mobility related information, $c_{T,M}$ represents the relative importance of Node Utilisation related information to Mobility related information and $c_{T,L}$ indicates the relative importance of Node Utilisation to the Link quality related information. These variables take values from the set $\{0.2, 0.33, 1, 3, 5\}$ according to Table 3.3 which reflects the relative importance between pairs of criteria, e.g., $c_{L,M}=5$ means that Link related information strongly more important than Mobility related information and $c_{M,L}=0.2$ vice versa. In other words, the comparison matrix has three independent variables, each with five possible values. Thus, giving a total of 125 different combinations. Each of these combinations result in a distinct set of $\{c_{L,M}, c_{T,M}, c_{T,L}\}$ parameters. The rest of $c_{i,j}$ parameters are set to 1. The main target is to find the combination
that results in the best performance for CLWPR, and then compute the corresponding \( f_X \) parameters. Figure 3.7 shows the effects of \( \{c_{L,M}, c_{\Gamma,M}, c_{\Gamma,L}\} \) parameters on PDR and E2ED using statistical analysis. It can be seen that there is a correlation between the selected weights and the performance of the protocol. First of all, we can observe some clear trends in Figure 3.7(a)-(b) and (d)-(e) where PDR and E2ED are at the highest and lowest, respectively, for small values of \( c_{L,M} \) and \( c_{\Gamma,M} \). This suggests that \( c_{L,M} \) and \( c_{\Gamma,M} \) should be relatively small, meaning that link and utilisation related information are less important than mobility. On the other hand, in Figure 3.7(c), PDR trend suggest that a relatively medium/high value for \( c_{\Gamma,L} \) provides better results. With these considerations in mind, we select the combination \( \{0.2, 0.2, 3\} \) for the relative importance coefficients \( \{c_{L,M}, c_{\Gamma,M}, c_{\Gamma,L}\} \). This set suggests that Mobility related information is strongly more important than Link quality related information \( (c_{L,M}=0.2) \) and strongly more important than Node Utilisation related information \( (c_{\Gamma,M}=0.2) \). Furthermore, Link quality related information is moderately less important as Node Utilisation related information \( (c_{\Gamma,L}=3) \).

Next, we evaluate the impacts of mobility related parameters \( f_D \), \( f_R \) and \( f_P \) in (3.4). The comparison matrix for these parameters is:

\[
\begin{bmatrix}
1 & c_{D,P} & c_{D,R} \\
 c_{P,D} & 1 & c_{P,R} \\
 c_{R,D} & c_{R,P} & 1
\end{bmatrix},
\]

(3.13)

where \( c_{P,D} \) represents the relative importance of Proximity related information to Distance, \( c_{R,D} \) represents the relative importance of Road related information to Distance and \( c_{R,P} \) the relative importance of Road related information over Proximity related information. Similarly to the previous evaluation, these variables can take values from the set \( \{0.2, 0.33, 1, 3, 5\} \), whereas the rest parameters are set to 1; thus, another set of 125 distinct combinations is formulated. The results of these simulations are presented in Figure 3.8 together with the statistical limits. In this set, the selection is not so clear as before; however, we considered the trade-off between PDR and E2ED metrics in our approach. In other words, there might not be an obvious difference for average PDR vs \( c_{P,D} \) (Figure 3.8(a)), but there is for E2ED (Figure 3.8(d)). Similarly
for $c_{R,D}$, there is a clear trend in PDR (Figure 3.8(b)) without significant difference in E2ED (Figure 3.8(e)). However, for $c_{R,P}$ both PDR and E2ED exhibit clear trends (Figure 3.8(c)-(f)), so the results can be highly objective. The corresponding selected set is $\{1, 5, 3\}$, which suggests that Proximity information is equally important as Distance related information ($c_{P,D}=1$), and moderate less important than Road related information ($c_{R,P}=3$). Road related information on the other hand is more important than Distance related information ($c_{R,D}=5$).

Following the same approach we evaluate the impact of link quality related factors, $f_C$,
3.6. Performance Evaluation of CLWPR

Figure 3.8: Performance of CLWPR protocol for different \( \{ c_{P,D}, c_{R,D}, c_{R,P} \} \) parameters

\[ f_{Ma}, \text{ and } f_{N} \text{ on (3.7). We formulate the comparison matrix:} \]

\[
\begin{bmatrix}
1 & c_{C,M_a} & c_{C,N} \\
 c_{M_a,C} & 1 & c_{M_a,N} \\
 c_{N,C} & c_{N,M_a} & 1
\end{bmatrix},
\]

(3.14)

where \( c_{M_a,C} \) represents the relative importance of MAC to CSI related information, \( c_{N,C} \) represents the relative importance of Neighbour Reliability to CSI related information and \( c_{N,M_a} \) is the relative importance of Neighbour Reliability to MAC related information. Keeping the other parameters to 1, we set each parameter to have a value from the set \{0.2, 0.33, 1, 3, 5\}. Based on the similar analysis of the observed trends for PDR and E2ED in Figure 3.9 we conclude that the optimal configuration set is \{0.33, 0.2, 0.2\}. This means that MAC related information is less important than CSI related
information \(c_{Ma,C}=0.33\) and moderate more important than Neighbour Reliability \(c_{N, Ma}=0.2\). CSI related information is moderate more important than Neighbour Reliability \(c_{N, C}=0.2\).

Finally, using the three sets for each group of relative importance parameters, we can calculate the set of optimal values for the \(f_X\) parameters using \(3.11\), as shown in Table 3.5. It is noted that the results in this subsection also demonstrate the effects of different parameters on the performance of the forwarding mechanism. This reveals an important shortcoming of the existing works presented in Section 2.3.1; they only consider mobility related information in the forwarding mechanism. Mobility is found to be the most important parameter, however others play non-negligible role. In addition,
3.6. Performance Evaluation of CLWPR

Table 3.5: Optimal Values for CLWPR parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_M$</td>
<td>0.0897</td>
</tr>
<tr>
<td>$f_L$</td>
<td>0.6070</td>
</tr>
<tr>
<td>$f_T$</td>
<td>0.3033</td>
</tr>
<tr>
<td>$f_D$</td>
<td>0.4796</td>
</tr>
<tr>
<td>$f_P$</td>
<td>0.4055</td>
</tr>
<tr>
<td>$f_R$</td>
<td>0.1150</td>
</tr>
<tr>
<td>$f_C$</td>
<td>0.1019</td>
</tr>
<tr>
<td>$f_{Ma}$</td>
<td>0.2121</td>
</tr>
<tr>
<td>$f_N$</td>
<td>0.6860</td>
</tr>
</tbody>
</table>

Our simulation results indicate that when a realistic propagation model is considered, link quality related information such as SINR becomes important, similarly to mobility related information. Therefore, for efficient next-hop selection, it is not sufficient to consider only mobility information; but link quality and node utilisation is also needed.

3.6.2 Comparison with ETSI-GF, AGF and GyTAR

With the optimal values for CLWPR identified in the previous subsection, we compare its performance against ETSI-GF, AGF and GyTAR. These protocols are proposed and evaluated in similar vehicular scenarios. Our simulation scenarios comprise of low, medium and high vehicle densities of approximately 5, 10 and 20 veh/km, respectively.

Scenario 1 (Manhattan grid scenario with synthetic traffic): Figures 3.10 and 3.11 compare the performance of the aforementioned protocols in terms of PDR, E2ED and overhead for V2V and V2I connections. CLWPR is shown to outperform the other protocols with respect to PDR (up to $\sim$20%) and overhead (up to 50%), while it has the lowest E2ED among protocols employing carry-n-forward (up to $\sim$10%). While ETSI-GF has the lowest E2ED since it does not cache packets, it also has the lowest PDR and highest overhead compared to the other protocols. AGF takes advantage of carry-n-forward mechanism and position prediction to increase PDR, which results in relatively high E2ED. On the other hand, GyTAR is able to achieve higher PDR, lower E2ED and lower overhead than AGF due to the use of traffic information. However, it is the combined criteria used by CLWPR protocol that assist to further increase PDR (up
to \( \sim 10\% \) and reduce the negative effect of carry-n-forward mechanism on E2ED (up to \( \sim 10\% \)). In addition, cross-layer optimisation allows CLWPR to select more resilient nodes that reduce the probability of retransmissions, and reduce end-to-end delay. Finally, the dynamic “HELLO” message exchange employed in CLWPR significantly reduces overhead compared to fixed broadcast interval in the rest protocols (up to 50%). The performance of all protocols is slightly improved for V2I communications scenarios as shown in Figure 3.11. This is due to the presence of fixed points resulting in less frequent path changes compared with moving destinations.

**Scenario 2** (“Unterstass” city scenario with real traffic): This scenario simulates a large scale real city scenario (Figure 3.6), where the effect of local maximum problem is stronger. This is manifested by the relatively low PDR of ETSI-GF protocol and the relatively high E2ED in protocols with carry-n-forward mechanism (more cached packets). In addition, the average number of hops is significantly larger than those
in Scenario 1 due to the longer distance between the end nodes in the flows, which increases end-to-end delay. Furthermore, in this scenario the overhead is substantially increased due to the higher number of nodes. Nevertheless, the trends in the results are similar to Scenario 1; CLWPR exhibits the highest PDR (up to 40%) and lowest overhead in general (up to ~20%), while it has the lowest E2ED among the protocols with carry-n-forward mechanism.

3.7 Discussion and Summary

As presented in the previous section, we use an AHP approach to tune the important parameters of the proposed protocol. The outcome is twofold. First, the optimised CLWPR protocol demonstrate significant advantages in performance over ETSI-GF, AGF and GyTAR protocols. Secondly, the insights we learn about the effects different environmental and communication parameters have on the performance of routing. From the results, it is manifested that mobility related information is not the only parameter to be accounted for in forwarding. A cross-layer approach should be considered for efficient performance. The proposed framework is evaluated using a fixed set of parameters, optimally adjusted for the two scenarios. This has kept computation complexity low as demonstrated by the simulation time evaluation (Appendix A, Figure A.1c). However, dynamic $f_X$ parameters could be used depending on the situation and re-adjusted based on additional information learnt by the system,
such as traffic information. Nevertheless, this would incur to additional overhead and complexity, that the current framework aims to minimise. With respect to complexity, the proposed protocol is based only on local information learned through the neighbor discovery mechanism with relatively simple mathematical calculations. In addition, the use of carry-n-forward keeps the complexity low, compared to other recovery mechanism, such as perimeter forwarding that needs to calculate planarized graphs.

With respect to the security and privacy issues concerned to the inherent ad-hoc and broadcast nature of the protocol, CLWPR as well as most position based routing protocols, rely on location service to provide them with the position information of the destination. There is a possibility of malicious attacks which is related to this process, that could potentially undermine the system by providing false information; thus, diverting the traffic away from the destination. Additionally, broadcasting periodically “HELLO” messages imposes privacy concerns. Both of these issues are covered in principles by the security architecture proposed in [14] with the help of security authorities that provide and verify pseudonyms to the users, i.e., vehicles.

In summary, we have presented a comprehensive study of the performance of routing protocols in distributed vehicular networks and we have proposed a novel and efficient routing protocol for VANETs. It considers mobility and cross layer information from PHY and MAC layers, in a joint weighting function in order to make effective forwarding decisions. With the help of Analytic Hierarchy Process, we optimise the relative weight assignment of the weighting function components, in order to enhance the performance of CLWPR. The performance analysis suggests that mobility related information is not the only criteria that should be considered in making forwarding decisions; link quality and utilisation related information are also important and increase performance by $\sim 10\%$. The comparison of the proposed protocol with the greedy forwarding algorithm proposed by ETSI, an advanced greedy forwarding algorithm and GyTAR, shows that the carry-n-forward mechanism as well as a prediction policy can increase PDR (up to 40%) with the cost of increase in End-to-End Delay. However, the use of cross-layer information can reduce the impact of caching on delay (up to 10%) and further increase PDR. Finally, dynamic broadcast should be considered in order to cope with overhead, which can potentially be halved.
Chapter 4

Hybrid Network Location Service Architecture

4.1 Introduction

The Location Service (LS) architecture is based on the client-server (pull technology) paradigm with two main processes. The first process is the location update where a client sends its location information to one or more servers. The second process is the position query, where a node asks one or more servers about the location information of a destination node. Location Services for MANETs are well studied \cite{126,128} and the aim in these LS is mainly to reduce overhead introduced and to increase success rate of queries. The expected routing overhead for this kind of LS has been formulated in \cite{129} as $\Omega(n^{1.5} \log(n))$, where $n$ is the number of nodes, assuming the mobility of the nodes is independent. Such an assumption is not valid in vehicular environments where car-following models usually describe the mobility of the nodes. Since MANETs are usually infrastructure-less, LS design is also based on distribution of the service among the mobile nodes. However, in vehicular environments we can capitalise on the existence of infrastructure; either that of cellular networks or dedicated Roadside Units (RSUs). A more detailed analysis of different LS is presented in Section 2.3.2.

In this chapter, we propose and evaluate a centralised location service architecture that is based on the existence of infrastructure (Section 4.2). In order to off-load the
wireless IEEE 802.11p-based access network, we propose a hybrid solution that utilises also existing cellular network, e.g. LTE. Such capabilities are feasible for vehicles and are evaluated now in field trials in projects like DRIVE C2X [22]. The results of the performance evaluation presented in Section 4.3 suggest that in higher traffic loads and higher vehicle densities, homogeneous networks, e.g. IEEE 802.11p, LTE, suffer from congestion due to limited resources and inefficient resource management. The proposed hybrid network architecture however, can cope better in such scenarios as traffic is split among the two networks.

4.2 Evaluated Location Service Network Architectures

Intuitively, a LS utilising infrastructure support, which can be available in vehicular scenarios, can improve the performance of the system. A centralised location server might be seen as single point of failure. However, with the introduction of cloud computing, it can be realised as a cloud service that will be available over a specific address, thus increasing the reliability of the service and resistance to node failures. We propose and evaluate two network architectures for LS that employ infrastructure
and a centralised location server as seen in Figure 4.1; one working purely with short-range ad-hoc network, e.g. IEEE 802.11p-based, and one utilising cellular network, e.g. 3GPP LTE, for location service traffic and short-range ad-hoc for inter-vehicle data traffic. Each of these architectures will be described in subsequent Section 4.2.2 and 4.2.3, respectively.

4.2.1 Location Service Operation

Vehicles exchange frequent 1-hop HELLO messages as a means of neighbour discovery mechanism. In addition, LSUPDATE messages destined to the remote location server are transmitted, which can be triggered by a timer or the distance travelled by a vehicle; in this thesis, we consider the timer approach. When a vehicle sends a packet to a destination, it first looks up its own local register for location information of the destination vehicle, to start the forwarding process, as presented with the CLWPR algorithm in Figure 3.1. If the required information is not locally available, a vehicle sends a LSREQ message to the LS server requesting the location information of the destination. This process is also performed at intermediate hops unless the location information is piggybacked to the data packets as explained in Section 4.2.4. These messages are sent either through the nearest RSU or BS to the location server, which replies back with a LSREPLY message. The location information is then stored on the local register for a certain valid time period. The validity period is related to the LSUPDATE inter-arrival time and is calculated as 1.5x times of it. This ensures that a new LSUPDATE message is received before the previous entry is purged. A sequence diagram of location information update and request for two vehicles is shown in Figure 4.2. Here, Vehicle 2 has some data for Vehicle 1. Both vehicles have already send periodic LSUPDATE messages to the LS server, which has created the corresponding entries in its database. When Vehicle 2 first identifies it needs to send data to Vehicle 1 (point R1 in figure), it queries the LS server for that information. The server will reply with the LSREPLY message and then Vehicle 2 is able to send the data to Vehicle 1 using the underlying position-based routing protocol.

\[^2\text{Inter-vehicle communications and ordinary users are not depicted for clarity}\]
4.2.2 IEEE 802.11p-based Location Service Architecture

The reference scenario for this architecture can be seen in Figure 4.1a. In this scenario, inter-vehicle communications as well as vehicle-to-infrastructure are achieved over the IEEE 802.11p network. Assuming that location service packets are not forwarded from other vehicles, RSU deployment has to be very dense to cover every possible street. For example, in urban areas, due to channel characteristics with building blocking line-of-sight communications, RSUs should be placed at every intersection as depicted in the reference scenario. The IEEE 802.11p-based LS architecture is presented in Figure 4.3. Vehicles are equipped only with IEEE 802.11p network interfaces and are connected with RSUs and other vehicles through them. RSUs are connected with a backbone network to the internet and through this to the location server. Vehicles send unicast location updates (LSUPDATE) and queries (LSREQ/LSREPLY) to the location server that are routed through the nearest RSU. One drawback of this approach is the need of large number of RSUs to have ubiquitous coverage. If location service messages were to be forwarded by other vehicles, then the number of RSUs could be lowered. Additionally, the use of same channel for location update and data dissemination, increases the contention levels which has negative impacts on the performance of the system. The benefit of this approach, though, is that vehicles need only one type of transceiver.
4.2. Evaluated Location Service Network Architectures

4.2.3 Hybrid Network Location Service Architecture

The second network architecture is depicted in Figure 4.4, where inter-vehicle communications are performed over IEEE 802.11p links but LSUPDATE and LSREQ/LSREPLY are routed through existing cellular network (e.g., 3GPP LTE). The benefits of such an approach are threefold: (a) utilising existing cellular infrastructure and not requiring dedicated RSUs, (b) the communication range of LTE is larger than IEEE 802.11p, thus fewer base stations are required for covering larger areas, and (c) we offload IEEE 802.11p network from the overhead introduced by LS. However, vehicles are required to have two types of network interface cards and packets that pass through the LTE core are potentially experiencing more delay.

4.2.4 Piggybacking Location Header

In addition to the two proposed network architectures for LS, the effect of piggybacking location information in the form of Location Header (LH) to data packets is investigated. With this technique, only the source node will have to query the LS for location information of the destination node. When it sends a packet, it piggybacks that in-
formation to the packet, so intermediate nodes shall not have to send queries to \( LS \). This approach potentially reduces the delay and overhead introduced by the \( LS \), but decreases the goodput\(^3\) of the wireless communications.

### 4.3 Performance Analysis of Location Service

In this section, we present the performance evaluation of the two proposed LS network architectures (IEEE 802.11p-based and Hybrid), with and without the use of Location Header, along with a full LTE network where all data are routed through the cellular network. Background traffic is generated in the LTE access network as well as the Internet for more realistic scenarios. The simulation area is a 5x5 Manhattan network; a benchmark scenario in the literature \[124\]. We simulated scenarios of different vehicle traffic density, vehicle speed and offered load using NS-3. A summary of the simulation parameters are presented in Table \[4.1\]. The performance metrics we used are the average end-to-end delay of data packets, and the overhead introduced by the location service (ratio between LSREQ sent and received packets). In addition, we evaluated the success

\(^3\)Ratio of useful information over total information sent.
4.3. Performance Analysis of Location Service

Table 4.1: Simulation Parameters for Location Service evaluation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of vehicles</td>
<td>100, 200, 300, 400</td>
</tr>
<tr>
<td>Vehicle Avg. Velocity</td>
<td>0 - 20m/s</td>
</tr>
<tr>
<td>Number of RSUs / eNB</td>
<td>25 RSUs / 4-9 eNB</td>
</tr>
<tr>
<td>Nominal Comm. Range</td>
<td>500m (at line-of-sight) &amp; shadowing</td>
</tr>
<tr>
<td>Number/Type of Connections</td>
<td>10 / UDP, V2V</td>
</tr>
<tr>
<td>Offered Load (per connection)</td>
<td>1-20 KBps / 500bytes/packet</td>
</tr>
<tr>
<td>MAC/PHY protocol</td>
<td>IEEE 802.11p, 6Mbps</td>
</tr>
<tr>
<td>LTE scheduler / RB alloc.</td>
<td>Prop. Fairness / 75 UL, 100 DL</td>
</tr>
<tr>
<td>VANET Routing protocol</td>
<td>CLWPR, cache limit 5sec</td>
</tr>
<tr>
<td>HELLO interval</td>
<td>Adaptive with speed</td>
</tr>
<tr>
<td>Loc. Service Update interval</td>
<td>5sec (time triggered)</td>
</tr>
<tr>
<td>Background Traffic</td>
<td>15 uE-uE connections (64kbps/con)</td>
</tr>
<tr>
<td>Internet Delay / Traffic</td>
<td>average 25ms / 50% link utilisation</td>
</tr>
</tbody>
</table>

ratio (LSREPLY / LSREQ) of the Location Service under different request rates. All the results are averaged over 15 independent simulation runs.

4.3.1 Location Service Success Ratio

We measure the success ratio of the Location Service Requests of the two architectures (IEEE 802.11p and Hybrid). The requests are sent following Poisson distribution with different rates and randomly selected pairs. Each request is sent once, and if it fails it is not retransmitted. The results presented in Figure 4.5 correspond to a scenario with 100 vehicles, moving with average speed of 15m/s. As it is expected, the demand on the location service has an effect on the success rate. For lower demand, IEEE 802.11p-based architecture can provide almost 100% success rate due to the assumptions of ubiquitous coverage from the RSUs. However, the rate is degraded for higher demand due to increase contention level and packet losses due to collisions. We evaluated two scenarios for the hybrid network; with 4 and 9 eNBs in the reference area, respectively. For both of them, the LS success rate is not affected by the LS load in such an extend as in IEEE 802.11p-based network, since LS demand is much lower compared to background traffic.
4.3.2 End-to-End Delay

Next, we present how *end-to-end delay* is affected by different parameters such as average vehicle speed, node density and traffic load. The results presented in Figure 4.6a suggest that LTE-based networks are not affected by the average speed of the vehicles as much as IEEE 802.11p-based ones, due to the large coverage area of the cells. However, node density and traffic load influence LTE-based networks more, making it a less desirable choice (Figures 4.7a, 4.8a). Networks with 4 eNBs exhibit higher delay than those with 9 eNBs, due to more resources available causing lower contention for the channel. However, handover delays could potentially affect the end-to-end delay when large number of eNBs is used. In this work, we have assumed an ideal handover mechanism though, which does not introduce delay. IEEE 802.11p-based networks demonstrate low delays which are affected primarily by the network size and load due to increased contention levels. On the other hand, hybrid networks result in lower delay for the most challenging scenarios (high mobility, density and load), which is explained by the split of data and location service traffic, which leaves some spare capacity on the IEEE 802.11p network. The number of eNBs also plays a role in the hybrid network architecture, as it increases the delay of the LS traffic. The effect of piggybacking Location Header (scenarios with LH) on delay is more apparent in high mobility where
4.4 Summary

In this chapter, we proposed and evaluated two network architectures for centralised Location Service in urban VANET scenarios; a homogeneous IEEE 802.11p-based

Figure 4.6: Impact of vehicles’ speed on end-to-end delay and overhead

the lack of intermediate $LS$ requests reduces the delay. Due to the highly dynamic network in those scenarios, intermediate nodes change rapidly and without LH, all of them would have to query the LS server, which introduces delay.

4.3.3 Overhead

Then, we evaluate the overhead introduced by the Location Service. As expected, increasing the average vehicles’ speed and node density result in equivalent increase of the overhead (Figures 4.6b, 4.7b). This is due to the dynamic nature of the network and the interconnections. Different nodes are selected per hop, so more frequent requests are sent to the location server. It is expected that the use of LH reduces the overhead in higher velocities. On the other hand, overhead is decreased as the traffic load is increased (Figure 4.8b) due to the definition of overhead. During the location information validity period described in Section 4.2.1 more data can be received as traffic load is increased, hence overhead is reduced.

4.4 Summary
Figure 4.7: Impact of vehicles’ density on end-to-end delay and overhead

Figure 4.8: Impact of traffic load on end-to-end delay and overhead
and a heterogeneous combining IEEE 802.11p and LTE networks. The results suggest that in higher traffic loads and vehicle densities, congestion and capacity limit the performance of homogeneous networks. The use of LTE network only for traffic related to $LS$ off-loads some traffic from IEEE 802.11p network and does not introduce excessive load on the LTE network. For future infotainment ITS applications the use of pure LTE networks could be an option; however a large number of sites should be deployed (potentially femtocells), which increases the cost of infrastructure. In addition, LTE networks may not be dedicated to ITS services, there are other users that increase the background load on this network. Therefore, using dedicated IEEE 802.11p-based access networks to deliver data in VANETs seem more suitable.
Chapter 5

Delay Bound Modelling

5.1 Introduction

In the previous chapters we analysed the proposed routing protocol and the location services taking into account average end-to-end delay. This metric is useful to evaluate the performance of such networking architectures and protocol. However, in order to get more information about the delay characteristics we have to look at other statistical metrics. In this chapter we model the upper bound of end-to-end delay for location-based routing in vehicular networks. We use *Stochastic Network Calculus (SNC)* \cite{130} to model the upper bounds of the end-to-end delay for three different vehicular network architectures: (a) only a vehicular ad-hoc network based on a short range communication technology, (b) only a cellular network with a large coverage area, and (c) a hybrid network comprising an ad-hoc and a cellular network as introduced in previous chapter. The detailed network model for (a) and (c) was presented in the previous chapter (Section 4.2).

The remainder of this chapter is organised as follows. Section 5.2 provides an overview of SNC, the methodology, notations and theorems that will be used later in our model. Section 5.3 describes the formulation of the problem in terms of SNC methodology with corresponding subsections for the arrival processes, and delay bounds for the three aforementioned scenarios. Section 5.4 presents the validation of the model and performance evaluation of the three scenarios in different configurations.
5.2 Overview of Stochastic Network Calculus

Network Calculus (NC) is a framework to analyse queueing/flow systems used for modelling in communication networks. It originated from the work of Cruz [131], which introduced an alternative to the classical queueing theory for analysing backlog and delay in communication networks. It uses more relaxed characterisation of distributions, which are defined by violation probabilities of arrival and service processes. Thus, providing bounds for the delay and backlog compared to the exact analysis of queueing theory, which are not tractable for complex real systems. NC employs min+/max+ algebra, which can transform non-linear queueing systems into analytically tractable linear systems. There are two different branches on NC: a) the Deterministic Network Calculus [132], which provides a worst-case analysis, however the bounds might be too loose; b) the Stochastic Network Calculus (SNC) [130], which gives a stochastic analysis with tighter bounds on the expense of small violation probabilities.

5.2.1 Definitions and Notation

Consider a service system as shown in Figure 5.1 with input $A(t)$ and output $A^*(t)$ after a variable delay. There are the following definitions and notations in the Network Calculus framework [130, 132]:

- **Arrival Process $A(t)$**: the total cumulative number of bits or packets arrived on the input flow in the time interval $(0,t]$. In addition, $A(s,t) = A(t) - A(s), \forall s < t$.

- **Stochastic Arrival Curve - (SAC)**: a flow is constrained by a wide-sense increasing function $\alpha(t)$, if for all $s \leq t : A(s,t) \leq \alpha(t - s)$, where $\alpha(t)$ is the arrival curve.

![Figure 5.1: Basic Input-Output System](image-url)
Table 5.1: SAC for different arrival types

<table>
<thead>
<tr>
<th>Type</th>
<th>Arrival Curve $\alpha(t)$</th>
<th>Bounding Function $f(x)$</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant Inter-Arrival</td>
<td>$T \cdot L \cdot t$</td>
<td>0</td>
<td>- $T$ packet arrival interval</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- $L$ packet size</td>
</tr>
<tr>
<td>Poisson</td>
<td>$r \cdot t$</td>
<td>$1 - (1 - a) \sum_{i=0}^{k} \left[a(i-k)! \frac{a(i-k)}{i!} e^{-a(i-k)}\right]$</td>
<td>- $\lambda$ arrival rate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- $L$ packet size</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- $r &gt; \lambda L$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- $a = \lambda L/r$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- $k = \left\lceil \frac{x}{L} \right\rceil$</td>
</tr>
<tr>
<td>gSBB [133]</td>
<td>$\rho \cdot t$</td>
<td>$me^{-nx}$</td>
<td>- $\rho$ upper rate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- $m, n$ optimisation parameters</td>
</tr>
</tbody>
</table>

for flow $A(t)$. There are different models to describe SAC, but in this thesis we focus on the virtual-backlog-centric (v.b.c.) model. A flow has a v.b.c. SAC $\alpha(t)$ with bounding function $f(x)$ denoted as $A \sim_{vb} \langle f, \alpha \rangle$, if $\forall t, x \geq 0$

$$P\left\{ \sup_{0 \leq s \leq t} \{ A(s, t) - \alpha(t - s) \} > x \right\} \leq f(x), \quad (5.1)$$

where $\sup\{S\}$ is the supremum of a subset $S$ of a totally or partially ordered set $T$ is the least element of $T$ that is greater than or equal to all elements of $S$. A list of common arrival processes used in networking has SAC presented in Table 5.1 [130].

- Departure Process $A^*(t)$: the total cumulative number of bits or packets seen on the output flow in the time interval $(0, t]$.

- Service curve defines the lower bound on the service provided by a server. The system is said to provide to the input a deterministic service curve $\beta(t)$ if

$$A^*(t) \geq (A \otimes \beta)(t), \quad \forall t \geq 0. \quad (5.2)$$
Here, $\otimes$ denotes the (min,+) convolution of two functions $F(t)$ and $G(t)$ as follows

$$(F \otimes G)(t) = \inf_{0 \leq \tau \leq t} \{F(\tau) + G(t - \tau)\},$$

where $\inf\{S\}$ is the infimum of a subset $S$ of a partially ordered set $T$ that is the greatest element of $T$ that is less than or equal to all elements of $S$. A widely used service curve type is the latency-rate service curve represented by $\beta(t) = Rt + T$, where $R$ and $T$ are the rate and latency parameters defined by the service process $S(t)$. Service process itself is a function of an underlying scheduling scheme. Examples of different schedulers and their corresponding rate and latency parameters are given in Table 5.2. There are different server models for SNC, however we only present the weak stochastic curve and the stochastic service curve (SSC) models that are used in our later analysis. A server $S(t)$ provides a weak stochastic service curve $\beta(t)$ with bounding function $g(x)$, denoted by $S \sim_{ws} (g, \beta)$, if for all $t \geq 0$ and all $x \geq 0$

$$P\{ (A \otimes \beta)(t) - A^*(t) > x \} \leq g(x). \tag{5.3}$$

A server provides a stochastic service curve (SSC) $\beta(t)$ with bounding function $g_t(x)$, denoted by $S \sim_{sc} (g_t, \beta)$, if for all $t \geq 0$ and all $x \geq 0$

$$P\{ \sup_{0 \leq s \leq t} [A \otimes \beta(s) - A^*(s)] > x \} \leq g_t(x). \tag{5.4}$$

If a server provides to the input a weak stochastic service $S \sim_{ws} (g, \beta)$, it provides a stochastic service $S \sim_{sc} (g_t^0, \beta - \theta)$ with the same service curve $\beta(t)$ and bounding function $g_t^0(x)$ equal to:

$$g_t^0(x) = \left[ \frac{1}{\theta} \int_{x-\theta t}^t g(y) dy \right]_1, \tag{5.5}$$

which holds for all $t \geq 0$, $x \geq 0$ and $\theta > 0$, $[z]_1 \equiv \min\{z, 1\}$.

- The virtual delay: is the delay that would be experienced by a bit or packet arriving at time $t$ if all bits (packets) received before it are served before it and
5.2. Overview of Stochastic Network Calculus

Table 5.2: Latency and rate terms of schedulers \[130\]

<table>
<thead>
<tr>
<th>Scheduler</th>
<th>Latency</th>
<th>Rate</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ideal General Processor</td>
<td>0</td>
<td>$\frac{\phi_i}{\sum \phi_i} C$</td>
<td>$C$ capacity of server [\phi_i] weight parameter for the input [L_{\text{max}}] maximum packet size [Q] quantum size allocated to the input</td>
</tr>
<tr>
<td>First-In-First-Out</td>
<td>$\frac{L_{\text{max}}}{C}$</td>
<td>$C$</td>
<td></td>
</tr>
<tr>
<td>Strict Priority</td>
<td>$\frac{L_{\text{max}}}{C}$</td>
<td>$C$</td>
<td></td>
</tr>
<tr>
<td>Deficit Round Robin</td>
<td>$(3 \sum Q_i - 2 Q_i)/C$</td>
<td>$\frac{Q_i}{\sum Q_i} C$</td>
<td></td>
</tr>
</tbody>
</table>

is given by

$$d(t) = \inf \{\tau : A(t) \leq A^*(t + \tau)\}. \quad (5.6)$$

There are a number of important theorems in the literature that are often used in SNC. Here, we only introduce the relevant theorems that are used in this thesis (the proofs can be found in \[130\]).

**Theorem 1 (End-to-End Delay Bound)** Consider a system with an arrival flow characterised by the arrival curve $\alpha(t)$ with bounding function $f(x)$, and the service has a stochastic service curve $\beta(t)$ with bounding function $g(x)$, then the virtual delay $d(t)$ satisfies the inequality

$$P\{d(t) > h(\alpha(t) + x, b(t))\} \leq (f \otimes g)(x), \quad (5.7)$$

where $h(a, b)$ is the maximum horizontal distance between functions $a(t)$, $b(t)$ and is defined as

$$h(a, b) = \sup_{s \geq 0} \{\inf \{\tau \geq 0 : a(s) \leq b(s + \tau)\}\}.$$ 

**Theorem 2 (Flow Aggregation)** Consider $N$ flows with arrival processes $A_i(t) \forall i = 1 \ldots N$. Then the aggregated arrival flow equals to the sum of all flows.

$$A(t) = \sum_{i=1}^{N} A_i(t), \quad (5.8)$$

and if $\forall i \ A_i \sim_{vb} (f_i, \alpha_i)$, then $A \sim_{vb} (f, \alpha)$ where $f(x) = f_1 \otimes f_2 \otimes \cdots \otimes f_N(x)$ and $\alpha(t) = \sum_{i=1}^{N} \alpha_i(t)$. 

Theorem 3 (Systems in Tandem) If a flow is traversing a sequence of servers \( i = 1, \ldots, N \) with constant propagation delay between the servers, each offering a stochastic service curve \( S \sim \langle g_i, \beta_i \rangle \) with service \( \beta_i(t) \) with bounding function \( g_i(x) \), the total (network) service curve \( \beta(t) \) and bounding function \( g(x) \) are given by

\[
\beta(t) = (\beta_1 \otimes \beta_2 \otimes \cdots \otimes \beta_N)(t),
\]

\[
g(x) = (g_1 \otimes g_2 \otimes \cdots \otimes g_N)(x).
\]

Theorem 4 (Leftover Service) Consider a system with an aggregated arrival of \( A(t) \), consisting of two flows \( (A_1(t), A_2(t)) \) and a stochastic service curve \( S \sim \langle g, \beta \rangle \). If flow \( A_2(t) \) has a v.b.c. SAC, \( A_2 \sim \langle f_2, \alpha_2 \rangle \), then the system guarantees to flow \( A_1(t) \) a stochastic service curve characterised by

\[
\beta'_1(t) = \beta(t) - \alpha_{2, \theta}(t),
\]

\[
g'_1(x) = (g \otimes f_{2,t}^\theta)(x).
\]

where \( a_{2, \theta}(t) = a(t) + \theta t \) and \( f_{2,t}^\theta = \frac{1}{\theta} \int_{x-\theta t}^\infty f_2(y)\,dy \).

5.3 End-to-End Delay Bounds

In this section, we describe the upper bound models for the end-to-end delay of location-based traffic including both data and signalling traffic, in three network architectures based on: (i) only short range ad-hoc wireless communications (e.g. IEEE 802.11p), (ii) only long range cellular communications (e.g. 3GPP LTE), and finally (iii) a hybrid network where the short range ad-hoc network is used for data communications and long range cellular network is used for signalling.

End-to-end delay is the sum of the delay in different communication layers, in one or multiple hops, depending on the scenario and network architecture. The delay in
5.3. End-to-End Delay Bounds

Each hop can be broken down into a number of components. For the short range communications, based on the IEEE 802.11p technology, the main source of delay for each hop is considered to be the time spent contenting for the shared channel, as well as any queueing delay. In long range communications, based on the 3GPP LTE technology, the total delay is a combination of delays introduced by the radio access network as well as the delay in Evolved Packet Core (EPC). Processing delays are neglected from our model since they are generally very small in the range of some microseconds \[134\]. The delay for communication from each RSU or the EPC to the LS server is mainly governed by the internet delay.

We use SNC to analyse the end-to-end delay; thus, we first describe the arrival processes in \[5.3.1\] with stochastic curves. Subsequently, calculations of the service curves and delay bounds according to \[Theorem\[7\]\] for the three aforementioned network architectures are given in Subsections \[5.3.2\] \[5.3.3\] and \[5.3.4\] respectively.

5.3.1 Modelling of Arrival Processes

In our system model, data and signalling traffic is generated from different sources, called arrival process here. This arrival process can be modelled by one of the generic traffic types with the corresponding v.b.c. \(\text{SAC} \ A \sim vb \ \langle f, \alpha \rangle\) presented in Table 5.1.

- **Application Data Traffic** depends on the type of application e.g., internet access, location advertisement or other infotainment application. In general, the traffic generated from these applications can be characterised by a generalised stochastically bounded bursty (gSBB) model \[133\].

- **Neighbor Discovery Traffic** is produced by periodic broadcast of 1-hop HELLO messages. The interval period can be static or dynamic in order to control the network overhead. In this thesis we consider static interval; thus, the HELLO message traffic can be characterised by a constant inter-arrival period process.

- **Location Service Traffic** consists of periodic LSUPDATE messages and asynchronous LSREQ and LSREPLY messages. The LSUPDATE messages can be
triggered by a timer or by the mobility of a vehicle. We adopt the timer approach, so the traffic is characterised by a constant inter-arrival period process. The LSREQ and LSREPLY messages are linked to the application traffic as explained in the previous subsection; their traffic is therefore characterised by a Poisson process.

- **Background Traffic** in cellular networks has long-tailed characteristics that can be modelled with a gSBB model.

5.3.2 End-to-End Delay Bounds: Ad-Hoc Network Scenario

In this subsection, we obtain the upper bound of the end-to-end delay for a vehicular ad-hoc network architecture based on short range communication technology. We consider a single channel network interface similar to IEEE 802.11p for DSRC. Here, each node is modelled by a stochastic process $S(t)$, which comprises a FIFO buffer and the second stochastic process $\hat{S}(t)$ to model the access to the shared channel as shown in Figure 5.2. $\hat{S}(t)$ characterises the service experienced by a packet that is at head-of-line (HOL) until it is successfully transmitted, otherwise known as access model. In IEEE 802.11p access model, access delay is dictated by a multi-stage backoff process. We use the model developed in [110] to calculate the mean access delay of a packet at HOL, $\bar{t}_{serv}$, as follows:

$$\bar{t}_{serv} = \sum_{j=0}^{R} p^j \bar{t}_j,$$

(5.13)

[^1]: Calculations for $p, E(b_j), t_B, t_{TX}$ can be found in [110].
where $R$ is the maximum number of backoff states, and $p$ is the collision probability. $\bar{t}_j$ is the mean time a node stays at backoff stage $j$, which is given by

$$\bar{t}_j = E(b_j)t_B + t_{TX}, \quad (5.14)$$

where $E(b_j)$ represents the number of backoff slots at stage $j$, $t_B$ is the average length of a backoff slot and $t_{TX}$ is the average length of a transmission slot. Since $\bar{t}_{serv}$ accounts only for access delay, we also need to calculate the average queuing delay in the FIFO buffer, which according to Little’s theorem \[136\] is given by:

$$t_q = \frac{\mathbb{P}}{\lambda}. \quad (5.15)$$

Here $\lambda$ is the average arrival rate of the packets, and $\mathbb{P}$ is the average number of packets in the queue. For a M/M/1/K queue \[137\]:

$$\mathbb{P} = \left\{ \begin{array}{ll}
\frac{K}{\rho} & \rho = 1 \\
\frac{K+1}{1-\rho(K+1)} & \rho \neq 1
\end{array} \right., \quad (5.16)$$

where $K$ is the queue size, $\rho = \lambda/\mu$, and $\mu = 1/\bar{t}_{serv}$ (pkt/sec).

In terms of SNC, $S(t)$ is described by a stochastic server with a stochastic service curve $\beta(t)$ bounded by $g_t(x)$, $S \sim_{sc} \langle g_t, \beta \rangle$, which following the work in \[112\] and using the Chernoff bound and the Lemma 2.2 in \[138\], are calculated as follows, $\forall x \geq 0$, if it makes $0 \leq y < 1 - q$.

$$\beta(t) = (\bar{t}_{serv} + t_q)\lambda \cdot t, \quad (5.17)$$

$$g_t(x) = \left\{ \frac{K}{y} \right\}^y \left( \frac{1-y}{1-y} \right)^{1-y}^K, \quad (5.18)$$

where

$$q = \frac{\bar{t}_{serv} + t_q - t_s}{Rt_c + Kt_{serv} + \mathcal{B}t_s}, \quad y = \frac{x - K \cdot t_s}{K(Rt_c + Kt_{serv} + \mathcal{B}t_s)}. \quad (5.19)$$

Here, $t_s$ is the average time the channel is busy due to successful transmission; $t_c$ is the average time the channel is busy due to collision; $R$ represents the maximum allowed number of retransmissions; and $\mathcal{B}$ is the maximum sum of backoff intervals given by.
\[ \sum_{r=0}^{\infty} (CW_r - 1), \text{ where } CW_r \text{ is the size of the contention window during backoff state } r. \]

The arrival process at each node \( i \), represents the aggregate traffic of data and signalling flows, i.e., HELLO messages and location service. Using Theorem 2 for aggregated SAC, \( A^i \sim_{vb} \langle f^i, \alpha^i \rangle \) is calculated as follows.

\[ \alpha^i(t) = \alpha^i_D(t) + \alpha^i_{LS}(t) + \alpha^i_H(t), \quad (5.20) \]

\[ f^i(x) = (f^i_D \otimes f^i_{LS} \otimes f^i_H)(x), \quad (5.21) \]

where \((\alpha_D, f_D), (\alpha_{LS}, f_{LS}), \text{ and } (\alpha_H, f_H)\) are the arrival curve and bounding function for data flow, location service flow, and HELLO messages, respectively.

Using Theorem 4 for leftover service, the service that each flow receives on node \( i \) can be calculated. For example, the service received by data flow, \( S^i_D \sim_{sc} \langle g^i_D, \beta^i_D \rangle \), is given by

\[ \beta^i_D(t) = \beta^i(t) - [\alpha^i_{LS,\theta}(t) + \alpha^i_{H,\theta}(t)], \quad (5.22) \]

\[ g^i_D(x) = (g^i_1 \otimes g^i_{LS,t} \otimes g^i_{H,t})(x), \quad (5.23) \]

where \( \alpha^i_{LS,\theta}(t) = \alpha^i_{LS}(t) + \theta t \), \( \alpha^i_{H,\theta}(t) = \alpha^i_H(t) + \theta t \), \( f^i_{LS,t} = \frac{1}{\theta} \int_{x-\theta t}^{\infty} f_{LS}(y)dy \) and \( f^i_{H,t} = \frac{1}{\theta} \int_{x-\theta t}^{\infty} f_{H}(y)dy \).

Further, based on the Theorem 3 for systems in tandem, the service that a flow will experience after \( n \) nodes is \( S^{net}_D \sim_{sc} \langle g^{net}_D, \beta^{net}_D \rangle \), where

\[ \beta^{net}_D(t) = (\beta^1_D \otimes \beta^2_D \otimes \cdots \otimes \beta^n_D)(t), \quad (5.24) \]

\[ g^{net}_D(x) = (g^1_D \otimes g^2_D \otimes \cdots \otimes g^n_D)(x). \quad (5.25) \]
The end-to-end delay bound for the data flow is given by Theorem 1 as

\[ P\{D_D > h(\alpha_D(t) + x, \beta_D^D(t))\} \leq (f_D \otimes g_D^D)(x). \]  

(5.26)

The Location Service traffic is routed from a vehicle, through a RSU towards the internet in order to reach the Location Server. In a similar way to that of the data traffic in (5.24) and (5.25), we calculate the service curve and bounding function for the LS traffic, \( S_{LS}^{net} \sim_{sc} \langle g_{LS}^{net}, \beta_{LS}^{net} \rangle \), as follows.

\[ \beta_{LS}^{net}(t) = (\beta_{LS}^1 \otimes \beta_{LS}^{int})(t), \quad (5.27) \]

\[ g_{LS}^{net}(x) = (g_{LS}^1 \otimes g_{LS}^{int})(x), \quad (5.28) \]

where \( S_{LS}^1 \sim_{sc} \langle g_{LS}^1, \beta_{LS}^1 \rangle \) is the stochastic service curve of wireless node for one hop to reach the RSU, and \( S^{int} \sim_{sc} \langle g_{LS}^{int}, \beta_{LS}^{int} \rangle \) is the stochastic service curve provided by the internet. Note that the internet is considered as a set of routers in tandem providing a latency-rate service with constant rate and a strict priority scheduling modelled by the service curve presented in Table 5.2 [120]. Thus, the end-to-end delay bound for the Location Service flow is given by

\[ P\{D_{LS} > h(\alpha_{LS}(t) + x, \beta_{LS}^{net}(t))\} \leq (f_{LS} \otimes g_{LS}^{net})(x). \]  

(5.29)

5.3.3 End-to-End Delay Bounds: Cellular Network Scenario

In this subsection, we examine the end-to-end delay model in a scenario with only a cellular network. We base our analysis on [120] and examine the delay bounds of the LTE network using the reference scenario of Fig. 5.3.

The arrival process \( A(t) \) in this network is the aggregation of two types of flow: a) background traffic (\( bg \)) which is assumed to consume almost 70-80% of system capacity, and b) the data traffic from the vehicles (\( veh \)). The traffic is characterised a \( v.b.c. \] SAC \( A \sim_{vb} \langle f, \alpha \rangle \) which is the aggregation of background and vehicle flows calculated using
To characterise the service curve for the LTE network, we examine the path that a
packet follows for both traffic flows. In the cellular only network scenario, a node sends
the packet through the ingress eNB to the EPC and the egress eNB, where it is delivered
to the destination node. The service curve of this system is the concatenation of three
subsystems: uplink, EPC and downlink. The uplink and downlink are governed by the
channel characteristics and the scheduler, while the EPC by the underlying network
capabilities of the core servers. This can be seen this as three systems in tandem,
therefore the total service curve, \( S \sim \langle \beta_{\text{net}}, g_{\text{net}} \rangle \), provided by the LTE network can
be calculated using Theorem 3 as follows.

\[
\beta_{\text{net}}(t) = (\beta_{\text{uplink}} \otimes \beta_{\text{EPC}} \otimes \beta_{\text{downlink}})(t),
\]

\[
g_{\text{net}}(x) = (g_{\text{uplink}} \otimes g_{\text{EPC}} \otimes g_{\text{downlink}})(x).
\]
In this thesis, we consider a SISO air interface for uplink and downlink with Round Robin scheduler, where the channel can be modelled as a two state Markov model; (i) ON state where transmission succeeds with probability of 1; (ii) OFF where a transmitted frame fails with probability of 1. The transition probability matrix is denoted by:

$$Q = \begin{bmatrix} q_{00} & q_{01} \\ q_{10} & q_{11} \end{bmatrix}$$

where $q_{ij} \in \{0,1\}$ denotes the transition probability from state $i$ to state $j$. The transition probabilities are calculated based on the system capacity of the LTE network. Assuming a system bandwidth of 10MHz, 1/3 coding rate and 16 QAM modulation scheme, the transmission rate is $\sim 11$Mbps. Now, considering the ON-OFF states of the Markov model, the ON state is set at 110Mbps and the OFF at 0. Therefore, the state transition probability $q_{01} = 1$ and $q_{10} = 0.1$ indicating a relatively fast fading speed [119].

The service curve of the channel is shown to have a stochastic service curve $\langle \beta(t), g(x) \rangle$ [139,140] with

$$\beta(t) = -\frac{1}{\theta} \log \frac{\omega(\theta)}{2} t,$$

$$\omega(\theta) = q_{00} + q_{11} e^{-c\theta} + \sqrt{(q_{00} + q_{11} e^{-c\theta})^2 - 4(q_{00} + q_{11} - 1)e^{-c\theta}},$$

$$g(x) = e^{-\theta x},$$

where $\theta$ is optimisation parameter, $c$ is the number of arrivals at the state ON. Selection of $\theta$ depends on the constraints for each specific traffic type. Figures 5.4 and 5.5 show the effect of $\theta$ on bounding function and service curve when $c = 1$. We will select the $\theta$ value that best fits our system. This depends on the constraints for each specific traffic as described in Table 5.3. In this system model, re-transmission until success is employed, which means no packet is dropped because of collision or deep channel fading. Packet losses only happen when the sojourn delay exceeds the delay budget. This gives us the following constraint for the LS traffic which is considered as signalling

$$P(Delay > 100ms) \leq 10^{-6}.$$
Table 5.3: Standardized QCIs for LTE

<table>
<thead>
<tr>
<th>QCI</th>
<th>Resource type</th>
<th>Priority</th>
<th>Packet delay budget (ms)</th>
<th>Packet error loss rate</th>
<th>Example services</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GBR</td>
<td>2</td>
<td>100</td>
<td>$10^{-2}$</td>
<td>Conversational voice</td>
</tr>
<tr>
<td>2</td>
<td>GBR</td>
<td>4</td>
<td>150</td>
<td>$10^{-3}$</td>
<td>Conversational video (live streaming)</td>
</tr>
<tr>
<td>3</td>
<td>GBR</td>
<td>5</td>
<td>300</td>
<td>$10^{-6}$</td>
<td>Non-conversational video (buffered streaming)</td>
</tr>
<tr>
<td>4</td>
<td>GBR</td>
<td>3</td>
<td>50</td>
<td>$10^{-3}$</td>
<td>Real-time gaming</td>
</tr>
<tr>
<td>5</td>
<td>Non-GBR</td>
<td>1</td>
<td>100</td>
<td>$10^{-6}$</td>
<td>IMS signaling</td>
</tr>
<tr>
<td>6</td>
<td>Non-GBR</td>
<td>7</td>
<td>100</td>
<td>$10^{-3}$</td>
<td>Voice, video (live streaming), interactive gaming</td>
</tr>
<tr>
<td>7</td>
<td>Non-GBR</td>
<td>6</td>
<td>300</td>
<td>$10^{-6}$</td>
<td>Video (buffered streaming)</td>
</tr>
<tr>
<td>8</td>
<td>Non-GBR</td>
<td>8</td>
<td>300</td>
<td>$10^{-6}$</td>
<td>TCP-based (for example, WWW, e-mail), chat, FTP, p2p file sharing, progressive video and others</td>
</tr>
<tr>
<td>9</td>
<td>Non-GBR</td>
<td>9</td>
<td>300</td>
<td>$10^{-6}$</td>
<td></td>
</tr>
</tbody>
</table>

bound on the arrival rate $\lambda$ of input traffic, which is defined as the capacity limit, $C = \max\{\lambda, \text{subject to Constraint}\}$.

In this system model, a re-transmission until success policy is employed, which means no packet is dropped because of collision or deep channel fading. Packet losses only happen when the sojourn delay exceeds the delay budget. $EPC$ is considered as a set of routers in tandem with constant rate and a strict priority scheduling modelled by the service curve presented in Table 5.2. Since the service is divided among the two traffic flows, the vehicle flow will get a fraction of server capacity according to Theorem 4 and is calculated as follows.

$$\beta_{veh}(t) = \beta_{net}(t) - \alpha_{bg,\theta}(t), \quad (5.36)$$

$$g_{veh}(x) = (g_{net} \otimes f_{bg,t}^\theta)(x). \quad (5.37)$$

where $\alpha_{bg,\theta} = \alpha_{bg} + \theta t$, $f_{bg,t}^\theta = \frac{1}{\theta} \int_{x-\theta t}^\infty f_{bg}(y) dy$. Finally, the delay of vehicle traffic in this case is bounded using Theorem 4 and is given by

$$P\{D_{veh} > h(\alpha_{veh}(t) + x, \beta_{veh}(t))\} \leq (f_{veh} \otimes g_{veh})(x). \quad (5.38)$$
5.3. End-to-End Delay Bounds

Figure 5.4: Bounding Function $g(x)$ for different $\theta$

Figure 5.5: Service Curve $\beta(x)$ for different $\theta$
5.3.4 End-to-End Delay Bounds: Hybrid Network Scenario

In the case of hybrid communications, each vehicle is equipped with two network interfaces; one for IEEE 802.11p and another for LTE. Data traffic and HELLO messages traffic are served by the IEEE 802.11p network, whereas Location Service flows and background traffic are served by the LTE network. According to Theorem 2, the arrival flows have v.b.c. SAC defined for the ad-hoc network by

\[ \alpha_{ah}(t) = \alpha_D(t) + \alpha_H(t), \] (5.39)

\[ f_{ah}(x) = (f_D \otimes f_H)(x), \] (5.40)

and for cellular by

\[ \alpha_{cell}(t) = \alpha_{bg}(t) + \alpha_{LS}(t), \] (5.41)

\[ f_{cell}(x) = (f_{bg} \otimes f_{LS})(x). \] (5.42)

The service curve, \( S_{ah} \sim_{sc} \langle g_{ah}, \beta_{ah} \rangle \), for the ad-hoc network, following the analysis in subsection 5.3.2, is given by

\[ \beta_{ah}(t) = (\beta_1 \otimes \beta_2 \otimes \cdots \otimes \beta_n)(t), \] (5.43)

\[ g_{ah}(x) = (f_1 \otimes f_2 \otimes \cdots \otimes f_n)(x), \] (5.44)

where \( n \) is the number of hops in the path of the flow and each service curve is calculated based on Theorem 4 for leftover service between the data and HELLO traffic. Thus, the delay bound in this case is given by

\[ P\{D_{ah} > h(\alpha_{ah}(t) + x, \beta_{ah}(t))\} \leq (f_{ah} \otimes g_{ah})(x). \] (5.45)

For the long range cellular communications the service is shared among the background
traffic and the Location Service traffic. The \(LS\) requests and updates are forwarded from the vehicles to the \(EPC\). From there, they pass through the internet towards the \(LS\) server; vice versa for the replies. Based on \textit{Theorem} 3, the service provided to the \(LS\) flows, \(S^{LS} \sim_{sc} \langle g^{LS}, \beta^{LS} \rangle\), is given by

\[
\beta^{LS}(t) = (\beta^{cell}_{LS} \otimes \beta^{int})(t),
\]

(5.46)

\[
g^{LS}(x) = (g^{cell}_{LS} \otimes g^{int})(x),
\]

(5.47)

where the \(\langle g^{cell}_{LS}, \beta^{cell}_{LS} \rangle\) is the characteristics of the service provided to the Location Service flow from the LTE network and \(\langle g^{int}, \beta^{int} \rangle\) is the characteristics of the service provided by the internet. According to \textit{Theorem} 4, the service left for the \(LS\) traffic in the LTE network is calculated as

\[
\beta^{cell}_{LS}(t) = \beta^{cell}(t) - \alpha_{bg, \theta}(t),
\]

(5.48)

\[
g^{cell}_{LS}(x) = (g^{cell} \otimes f^{\theta}_{bg,t})(x),
\]

(5.49)

where \(\alpha_{bg, \theta} = \alpha_{bg} + \theta t\), \(f^{\theta}_{bg,t} = \frac{1}{\theta} \int_{x-\theta t}^{\infty} f_{bg}(y) dy\) and \(\langle g^{cell}, \beta^{cell} \rangle\) is calculated for the uplink as

\[
\beta^{cell}(t) = (\beta^{uplink} \otimes \beta^{EPC})(t),
\]

(5.50)

\[
g^{cell}(x) = (g^{uplink} \otimes g^{EPC})(x).
\]

(5.51)

and for the downlink as

\[
\beta^{cell}(t) = (\beta^{downlink} \otimes \beta^{EPC})(t),
\]

(5.52)

\[
g^{cell}(x) = (g^{downlink} \otimes g^{EPC})(x).
\]

(5.53)

Finally, the delay bound of the Location Service flow in the hybrid network is given by
5.4 SNC Model Validation and Performance Evaluation

In this section, we evaluate the models for the three scenarios and validate them through simulations using the NS-3 simulator (Subsection 5.4.1). In addition, we perform extensive performance evaluation of the aforementioned network architectures in terms of end-to-end delay (upper bounds and characteristics) for data and signalling traffic, and throughput (Subsection 5.4.2). Our reference area is depicted in Figure 5.6, where vehicles, ordinary cellular users, RSUs, BS and backhaul network are represented. Each vehicle can communicate directly to other vehicles in the same group and adjacent groups. Each RSU serves only one group, while only one BS serves all vehicles and all ordinary users of the LTE network in the reference area. Vehicles are equipped with IEEE 802.11p and/or LTE communication modules. The distance between each RSU is 300m, which is also the nominal communication range of the short range wireless modules of the vehicles. Vehicles travel with a constant speed (50km/h), while ordinary users are randomly distributed in the area with fixed positions. The configuration parameters for the simulations and the analysis are summarised in Table 5.4.

\[ P\{D_{LS} > h(\alpha_{LS} + x, \beta_{LS})\} \leq f_{LS} \otimes g_{LS}^{cell}. \] (5.54)
Table 5.4: Configuration Parameters for SNC evaluation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of vehicles</td>
<td>10 / 20 per group</td>
</tr>
<tr>
<td>Number of other users</td>
<td>100 random allocation</td>
</tr>
<tr>
<td>Data Packet Size</td>
<td>500Bytes</td>
</tr>
<tr>
<td>802.11p Data Rate</td>
<td>6 &amp; 27Mbps</td>
</tr>
<tr>
<td>Buffer size (Φ)</td>
<td>100 packets</td>
</tr>
<tr>
<td>LTE scheduler / RB alloc.</td>
<td>Round Robin / 25 RBs</td>
</tr>
<tr>
<td>Loc. Service Update interval</td>
<td>5sec (time triggered)</td>
</tr>
<tr>
<td>Data Traffic</td>
<td>V2V connections (10-20kbps/con)</td>
</tr>
<tr>
<td>Background Traffic</td>
<td>80 uE-uE connections (200kbps/con)</td>
</tr>
<tr>
<td>Internet Delay</td>
<td>average 10ms</td>
</tr>
</tbody>
</table>

5.4.1 Model Validation

Our first scenario consists of 10 vehicles per group and vehicles from group #1 send data to vehicles from group #4, forming a 3-hop communication at the IEEE 802.11p network, using 6Mbps data rate. Figures 5.7 and 5.8 show the numerical evaluation of the models and simulation results, and as it can be observed there is a relative tight approximation of the delay bound. For this scenario, the IEEE 802.11p-based and 3GPP LTE networks are closely competing with each other, while the proposed hybrid shows significantly lower bounds. In terms of location service traffic (Figure 5.8), the hybrid network can provide very low bounds compared to the IEEE 802.11p network. The LTE scheduler can provide stricter QoS as opposed to the enhanced distributed channel access (EDCA) mechanism of IEEE 802.11p. We can observe, the IEEE 802.11p curve has a very long tail. The second scenario evaluates IEEE 802.11p and hybrid networks when different number of hops are required to reach the destination. The LTE network is not affected by the number of hops, thus is not analysed in this scenario. For 2-hop communications, vehicles from group #1 send data to group #3 and for 3-hops to group #4. Figures 5.9a and 5.9b present the delay distribution over the IEEE 802.11p and hybrid network, respectively. It is obvious that more hops result in increased end-to-end delay, but even with only two hops, the proposed hybrid scenario can provide lower bounds compared to the IEEE 802.11p network. The third scenario evaluates how IEEE 802.11p data rate affects the delay bounds for pure ad-hoc and hybrid networks. Figures 5.10a and 5.10b present the delay distribution for the two architectures. As
it can be seen, the bound of pure ad-hoc network is decreased significantly compared to the decrease of hybrid network. This is due to the increase in capacity of the IEEE 802.11p network which benefits both data and signalling traffic for the pure ad-hoc network. However, the hybrid system is benefited only partially, as location service traffic is carried from the LTE network and is subject to the same delay in both configurations. We can observe that in the 27 Mbps scenarios, the two architectures provide relatively similar delay bounds. The available data rate is adequate to carry the data and location service traffic (for the IEEE 802.11p network) without increasing the collision probability, thus keeping the access delay low. In all evaluated scenarios, the proposed bound models provide a relative close fit to the simulation results.

5.4.2 Performance Evaluation

In the previous subsection, we validated the proposed analytical models and showed that the upper bounds calculated from the SNC methodology are relative tight. In this subsection, we evaluate the three network architectures, using both the proposed models and simulations, in different scenarios.
5.4. SNC Model Validation and Performance Evaluation

Figure 5.8: Comparison of model bound and simulation results for Loc. Service packets (10-nodes/group scenario)

Figure 5.9: Comparison of model bound and simulation results for different hop count

(a) IEEE 802.11p

(b) Hybrid
We increase the number of vehicles to 20 per group, and evaluate the 3-hop scenario, with the results presented in Figure 5.11. It is clear that the increase of vehicles affects the contention on the shared channel of IEEE 802.11p and the hybrid network, while the pure LTE is less affected because the proportion of contenting nodes does not increase in the same way. In this scenario, the IEEE 802.11p network delay bound is increased significant and now the hybrid network is closely competing with the LTE network. However, the results suggest that the hybrid network still can deliver better end-to-end delay (average and 75th-percentile) than the other two network architectures (Figures 5.12a and 5.12b).

In addition to the data traffic, signalling is also important. We evaluate the average end-to-end delay of Location Service traffic in IEEE 802.11p and the hybrid networks, accounting both for uplink and downlink flows. The results presented in Figure 5.13 suggest that the pure IEEE 802.11p networks provides lower delay, when the contention levels are low. However, the hybrid network is affected less by the increase in the number vehicles per group and can deliver lower end-to-end delay.

We evaluate the aforementioned network architectures in terms of average normalised throughput on the data flows (Figure 5.14). In the scenarios with lower contention on the shared channel (10 veh./group), hybrid architecture provides the highest normalised throughput compared to the other two architectures. Still, with the increase in the
5.4. SNC Model Validation and Performance Evaluation

Figure 5.11: Comparison of model and simulation results for data packets (3-hop & 20veh. scenario)

Figure 5.12: Comparison of end-to-end delay distribution for data traffic
number of vehicles per group, the limited available bandwidth of the IEEE 802.11p, when using 6Mbps data rate, reduces the achievable throughput in the pure 802.11p and hybrid architectures. This reduction is more evident in the 3-hop scenario, where the contention levels are increased. However, if we increase IEEE 802.11p data rate to 27Mbps, there is enough capacity to accommodate both data and signalling, so normalised throughout stays relatively higher than the LTE network.

5.5 Summary

An analytical model for the calculation of the end-to-end delay bounds for IEEE 802.11p, 3GPP LTE and hybrid vehicular networks is presented in this chapter. The use a Stochastic Network Calculus approach to transform the original problem into a mathematically tractable problem is proven to provide relatively tight upper bounds on the end-to-end delay for different networks that are considered in this work. The model considers the network as a whole, compared to hop-by-hop bound, thus its complexity is proportional to $\Theta(n \log n)$ as mentioned in Section 2.3.5. Different from existing results in the literature, we considered realistic multi-hop scenarios with unsaturated traffic models. The results of our investigation suggest that hybrid networks can sig-
5.5. Summary

significantly help improve the performance of vehicular networks, in terms of end-to-end delay bounds both for data traffic and signalling.

Figure 5.14: Comparison of throughput for different data rates
Chapter 6

Cost-Efficient Transport Protocol

6.1 Introduction

There is an ongoing debate in the research and industry community whether IEEE 802.11p or 3GPP LTE should be used for vehicular communications. As we demonstrated in the previous chapters, a single interface can not handle all traffic efficiently. Connected vehicles are promoted with the use of different communication technologies for diverse applications as we have already seen. A host with multiple network devices is referred to as a multi-homed node and Stream Control Transmission Protocol (SCTP) is an IETF standard which supports multi-homing. However, original SCTP multi-homing functionality is only used when the primary address becomes unavailable. There are several extensions to this protocol that exploit multiple network interfaces and paths available to increase throughput and reduce latency, which we review in section 2.3.3. However, their results are based on simulation evaluation, without providing a robust analytical model. In this chapter, we present an analytical model for a modified SCTP model (section 6.2), which selects the primary network interface based minimum round trip time (RTT) such as the proposals in [96, 98]. Furthermore, we propose a novel extension of SCTP that takes into account not only the path quality (e.g. RTT, available bandwidth), but also the cost of using individual links (section 6.4), based on the model described in section 6.3.
Chapter 6. Cost-Efficient Transport Protocol

Figure 6.1: Original SCTP Markov Chain states with multihoming

6.2 Analytical Model of RTT-aware SCTP

In this section we model the throughput of a modified SCTP protocol, which switches between primary and secondary paths not only based on time-out event as in the original SCTP specifications, but using a utility function, as presented in section 2.3.3, such as minimum RTT [96,98].

6.2.1 Original SCTP modelling

There are only two published works to the best of our knowledge in the literature that provide analytical model for SCTP throughput and are based on discrete Markov chain models. The work presented in [137] models SCTP with original multihoming functionality, where primary and secondary paths are alternated only at loss events as seen in Markov chain diagram in Figure 6.1. Each state has three elements \( \{cwnd, W_t, l\} \), where \( cwnd \) represents the congestion window size in segments, \( W_t \) represents the slow start threshold and \( l \) is an indicator of loss. The transitions are grouped in five categories as summarized below:\(^1\)

- **Slow Start**: from state \( \{w, W_t, 0\} \) to \( \{2w, W_t, 0\} \) with probability \( P_w(0) \),

---

\(^1\)For more information and notations on the equations please refer to [137]
6.2. Analytical Model of RTT-aware SCTP

- **Congestion Avoidance** from state \( \{w, W_t, 0\} \) to \( \{w+1, W_t, 0\} \) with probability \( P_w(0) \),

- **Time-out** from state \( \{w, W_t, 0\} \) to \( \{0, \lfloor w/2 \rfloor, 1\} \) with probability \( P_{TO}^w \),

- **Exponential Backoff** from state \( \{0, W_t, 1\} \) to \( \{0, 2, 1\} \) with probability \( P_1(1) \), and

- **Fast Retransmission** from state \( \{w, W_t, 0\} \) to \( \{\lfloor w/2 \rfloor, \lfloor w/2 \rfloor, 1\} \) with probability \( P_{FR}^w \).

This model calculates the expected number of segments generated per RTT as:

\[
G = \sum_{w=1}^{w_{max}} w P(\text{cwnd}^{(w)}),
\]  

(6.1)

where by solving the Markov model in steady state

\[
P(\text{cwnd}^{(w)}) = \sum_{W_t=2}^{w_{max}} \sum_{l=0}^{1} \pi(w, W_t, l),
\]  

(6.2)

where \( \text{cwnd}^{(w)} \) is equivalent to \( \text{cwnd} = w \) and \( w_{max} \) is the maximum available congestion window for that state. The expected lost segments from the primary path, i.e. traffic transferred into the secondary as:

\[
E[L] = \sum_{w=1}^{w_{max}} \sum_{k=1}^{w_{max}} k P(\text{loss}^{(k)} | \text{ccwnd}^{(w)}) P(\text{ccwnd} = w).
\]  

(6.3)

Here \( \text{ccwnd} \) represents the current congestion window.

The second work that provides an analytical model of SCTP is [141]. However, the proposed model does not consider the multi-homing functionality, while trying to provide higher accuracy in the steady-state throughput. It models the different states of a SCTP association, namely congestion avoidance (CA), exponential back-off (EB) after time-outs (TO) and slow-start (SS), and for each one, it estimates the number of packets and duration in steady state to calculate the throughput.

Nevertheless, in section 2.3.3, we presented SCTP enhancements, in which the selection of primary and secondary path is dynamic, based on some short of utility function.
The transition from primary to secondary is not triggered only by a TO/FR event, but also by the utility function (UF) event.

6.2.2 Modified SCTP model

In this section we present the modified SCTP model that takes into consideration the minimum RTT to select the primary path. Our work is based on the SCTP model in [137], however, we extend the state transition as shown in Figure 6.2 with the addition of an Utility Function (UF) event. The state transition for UF from \((w, W_t, 0)\) to \((w, W_t, 2)\) is with probability \(P_{UF}\). Our objective is to quantify \(P_{UF}\); the probability that according to the utility function\(^{1}\) there is a swap of primary and secondary path.

6.2.2.1 Definition of \(P_{UF}\)

According to our utility function, a UF event happens when the RTT of Network 1 is larger than that of Network 2 assuming that the current primary path is on Network 1, and vice versa. This is formulated as follows for \(x, y\) the two networks in [6.4], where

\[^{1}\text{We use min RTT in this work, but the model can be used in principle for other utility functions.}\]
6.2. Analytical Model of RTT-aware SCTP

\[ \theta \text{ is the RTT time.} \]

\[ P_{N_x}^{UF} = P\{\theta_{N_x} > \theta_{N_y} \mid N_x \text{ is primary}\}. \quad (6.4) \]

From Bayes formula we have:

\[ P\{\theta_{N_x} > \theta_{N_y} \mid N_x\} = \frac{P\{N_x \mid \theta_{N_x} > \theta_{N_y}\}P\{\theta_{N_x} > \theta_{N_y}\}}{P\{N_x\}}, \quad (6.5) \]

Now we need to estimate the RTT for each path. This results from the analysis of the network model as represented in Figure 6.3 and is the sum of queueing delay \((d_q)\) and a propagation delay \((d_t)\). Assuming that each individual network can be modelled as a M/M/1/K queue, similar to [137], with a fixed capacity \(BW\) and queue size \(K\), we can calculate its queueing delay using Little’s Law. Propagation delay can be assumed fixed and depended on the network technology used.

### 6.2.2.2 Queueing Delay Calculation

Assuming that each source \(i\) has a traffic flow \(\lambda_{i,k}\) on network \(\kappa\) and there are \(N\) and \(M\) flows on each network, the total traffic flow on each network is \(\lambda_{\kappa} = \sum_{i=0}^{k} \lambda_{i,\kappa}\) with \(k \in \{M, N\}\) and \(\kappa \in \{N_1, N_2\}\). Aggregation of a large number of SCTP traffic sources results in the overall traffic arrival to the network being Poisson. To find the average
queueing delay \( (d_{q,\kappa}) \) in a network we use Little’s Law and evaluate it as follows:

\[
d_{q,\kappa} = \frac{S_{\kappa}}{\lambda_{\kappa}},
\]

where \( S_{\kappa} \) is the amount of packets in the queue for network \( \kappa \) given by:

\[
S_{\kappa} = \begin{cases} 
\frac{K}{2} & \rho = 1 \\
\rho \left( \frac{K+1}{1-\rho^{K+1}} \right) & \rho \neq 1
\end{cases},
\]

and \( \rho = \lambda_{\kappa}/\mu_{\kappa} \), \( \mu_{\kappa} = (BW_{\kappa}/8) \text{PacketSize} \). Finally, RTT on Network 1 is calculated as \( \theta_{N_1} = d_{q,N_1} + d_{t,N_1} \). In a similar way, we can estimate the RTT of the second network, and finally find the probability \( \theta_{N_1} > \theta_{N_2} \).

### 6.2.3 Evaluation of RTT-aware SCTP model

In this section we evaluate the proposed model and compare it with \[137\] on throughput pre path. For the evaluation of the proposed model we used the following parameters: \( w_{\text{max}} = 32 \), capacity 100 MBps, queue size \( K=50 \), segment size 500 bytes, number of flows on each network 50, propagation delay 0.1 sec and we varied the flow rate \( \lambda \).

As we observe in Figure 6.4a and 6.4b, the throughput is decreased as the average flow rate is increased. This is due to congestion and increase of segment loss probability. In
addition, since the queueing delay is increased, the RTT is increased, which also has a negative impact on throughput. While the total throughput in the two cases (original and modified) remains the same, the distribution of traffic among the available paths is different. The modified SCTP utilises the secondary path more frequently as it can be seen in Figure 6.5. When the flow rate reaches the limit of the path capacity, both paths are equally utilised.

### 6.3 Cost Model

There are a lot of research works lately looking at mobile data offloading from an economic perspective rather than simply from a system’s performance point of view. Paolini in [142] and Dhawan et al. in [143] approach the problem using CAPEX and OPEX analysis of different systems, i.e. macro-cells, femto-cells, WiFi APs etc. These two works show how the total cost could be reduced by introducing more small cells and integrate WiFi with cellular in order to offload the later. They assume a fixed proportion of mobile data to be offloaded, e.g. 60% as reported in [144]. On the other hand, Gao et al. in [145] and Lee et al. in [146] provide analytical models based on

\(^{1}\text{CAPEX} = \text{capital expenditures, OPEX} = \text{operational expenditures}\)
game theory in order to find the equilibrium of offloading or the economic benefits of
offloading according to a certain pricing scheme.

We base our proposal for cost-efficient SCTP (presented in Section 6.4) on the findings
of the work in [146] focusing on the required offloading ratio in order to have economic
benefit for both consumers and provider. According to that work, users are modelled
with four attributes, (i) how much money they can pay (willingness to pay, $\gamma$), (ii)
how many data they want to use (traffic demand, $\phi$), (iii) how long their data can
tolerate (delay profile, $\alpha$), and (iv) how they move (WiFi contact probability, $e$). The
model assumes that the monopoly provider knows users’ attributes and strategies a
priori, and the market can be modelled based on a two-stage sequential game (e.g.
Stackelberg game). At the first stage, the provider decides on the pricing parameters
($p$) as a leader, and at the second stage, each user is a price-taker as a follower and
chooses its LTE+WiFi traffic volume $x$. The analysis results are carried out based on
the equilibrium of this game assuming $N$ total users and $\hat{N}$ users per cell. The user’s
net-utility is defined as $U(x)$ and the provider’s revenue as $R(p)$.

$$U(x) = \sum_{t \in T} \gamma(t)x^\theta(t) - m(p, y(x)),$$

(6.8)

where $\theta \in (0, 1)$ is the price sensitivity, $m(p, y(x))$ is the daily payment charge for usage
of LTE network, and

$$R(p) = \sum_{i \in N} m(p, y_i(x_i)) - \sum_{i \in N} c(y_i(x_i)),$$

(6.9)

where $c(y) = \eta \sum_{t \in T} y_i(t)$ is the network cost to handle the LTE traffic with $\eta$ the cost
per unit of data [147].

---

1We use “LTE” to refer to a cellular network in general.
6.3. Cost Model

6.3.1 Off-Loading indicator

The offloading indicator quantifies how much LTE data is offloaded, (i) aggregate LTE traffic ratio \( \kappa_{avg} \), and (ii) peak LTE traffic ratio \( \kappa_{peak} \).

\[
\kappa_{avg} = \frac{\sum_{t \in T} Y(t)}{\sum_{t \in T} X(t)},
\]

(6.10)

\[
\kappa_{peak} = \frac{\max_{t \in T} Y(t)}{\sum_{t \in T} X(t)},
\]

(6.11)

where the transmitted total traffic and LTE traffic over a cell at time \( t \), \( X(t) \), \( Y(t) \) are:

\[
X(t) = \hat{N} \int_{0}^{\Phi_{max}} x_{\Phi}(t) dF_{\Phi},
\]

(6.12)

\[
Y(t) = \hat{N} \int_{0}^{\Phi_{max}} \sum_{d=0}^{D} b_{\Phi}^{d}(t-d)x_{\Phi}(t-d) dF_{\Phi},
\]

(6.13)

and \( b_{\Phi}^{d}(t-d) = \alpha_{\Phi}^{d}(1 - e_{\Phi}^{d}(t)) \) is the portion of the traffic generated at time \( t \) which is transmitted through LTE at time \( t + d \). It is clear that as users delay more traffic, the aggregate LTE traffic ratio \( \kappa_{avg} \) provably decreases, since more traffic can be offloaded through WiFi.

Opt-saturated and opt-unsaturated are two defined notions, which characterize the regimes of how much traffic is imposed on the network for the equilibrium price. In general, as traffic demand gets higher compared to the LTE capacity, the network becomes opt-saturated, and vice versa. For a unique equilibrium price \( p^* \), the network is said to be opt-saturated if the network is saturated at \( p^* \), vice versa (Figure 6.6).

Theorem 3.1 in [146] states that for flat pricing\(^1\), if the cost of the unit volume of the LTE traffic \( \eta < (\kappa_{avg}\Phi_{max}^{1-\theta})^{-1} \) the net-utilities of all subscribers increase and the provider’s revenue at equilibrium increases as (i) \( \kappa_{peak} \) decreases in the opt-saturated case, and (ii) as \( \kappa_{avg} \) decreases in the opt-unsaturated case.

We assume that the price schemes available or those that will be available for vehicular customers, are in equilibrium. Therefore, the user of flat pricing should aim to have

\(^1\)The theorem is valid also for other pricing schemes
\[ \eta < (\kappa_{avg} \Phi_{max}^{1-\theta})^{-1} \]

so the net-utilities will increase according to Theorem 3.1 in [146], knowing the regime (opt-saturated or opt-unsaturated) that the system operates. For a fixed \( \eta, \theta \) and \( \Phi_{max} \) we can only control the offloading ratio. For example as shown in Figure 6.7 for a given \( \eta = 0.1\$/MB \), the maximum off-loading ratio for aggregate LTE traffic is 0.7 and 0.31 for a maximum daily demand of 200MB and 1GB assuming \( \theta = 0.5 \), while it lowered to 0.24 and 0.08 for \( \theta = 0.3 \), respectively. This shows the effect that both daily demand from the users and how much the users value that data have on the cost efficiency of mobile data offloading.

### 6.4 Cost-Efficient Transport Protocol

In the light of these findings, we propose a cost-effective transport protocol based on SCTP, named CE-SCTP and described by the switching algorithm in Figure 6.8, for hybrid vehicular networks. The algorithm has as inputs the system characteristics such as the traffic patterns, application requirements, and the cost model. These parameters will determine (a) the \( \kappa_{avg} \) limit in order to have a cost-efficient system and (b) the amount of data to be off-loaded in order to increase cost efficiency.

The selection of the best path is performed every RTT. QoS information that is used in our proposal is based on the estimated bandwidth for each path. The approach is similar to the work in [148], where the congestion control mechanism follows TCP-Westwood. Since only one path (the primary) is utilised per RTT\(^1\), we need to monitor

---

\(^1\)CMT approach not implemented at this stage
6.4. Cost-Efficient Transport Protocol

![Graph showing Maximum OffLoading Ratio Vs cost-coefficient](image)

**Figure 6.7: Off-Loading Ratio \( \kappa_{avg} \) for different traffic demands**

**Figure 6.8: Algorithm for proposed CE-SCTP**

1. **Traffic Patterns**: Initialize model
2. Calculate QoS information for every path
3. **Application Requirements**: Select list of paths that meet application requirements
4. **Cost model**: Evaluate net-utility for each path
5. Select path that is most cost-effective so 
   \[ \eta < (\kappa_{avg} \Phi_{max}^{1-\theta})^{-1} \]
the bandwidth of the other available paths. There are different methodologies that can be used for that purpose, such as Variable Packet Size (VPS) probing, Packet Pair/Train Dispersion (PPTD) probing, Self-Loading Periodic Streams (SLoPS), and Trains of Packet Pairs (TOPP), as reported in [149]. We use an approach similar to PPTD, where we send several probe packets and estimate available bandwidth using the same method as in primary path. Knowing the requirement of the application in terms of capacity and the average off-loading ratio in order to have a cost-efficient communication based on the previously described cost model, we select the primary path which best meets these two requirements. However, we need to take into account possible latency inequalities among the two paths. Therefore, the switching should not be performed very frequently but have some hysteresis in order to minimize possible jitter.

6.4.1 Architecture for CE-SCTP

Key elements of the CE-SCTP architecture are presented in a LTE-WiFi hybrid vehicular system as shown in Figure 6.9. The architecture borrows concepts from 3GPP LIPA/SIPTO [150], with the most important enhancement of the proposed architecture being the monitoring server. It performs several functionalities in a RESTful manner, interacting with the underlying cellular and wireless networks, as well as external hosts, which include:

- **Monitor data usage for both cellular and wireless networks.** In order to calculate the off-loading indicator, the mobile operator needs to keep track of all traffic in its network and off-loaded traffic through the wireless network.

- **Provide advice during a connection setup on $\kappa_{\text{avg}}$.** During a SCTP connection setup, the remote server\(^1\) will contact the monitoring server to get the most recent off-loading indicator for cost-efficient data transfers, based on the connection requirements and other service-level agreements (SLAs).

- **Provide interfaces to EPC.** In order to be able to monitor traffic and provide the advice, the monitoring server should provide interfaces to the EPC. These

\(^1\)Similar procedure happens from vehicle-initiated connections.
6.5 Performance Evaluation of Cost Effective Transport

In this section we present the performance evaluation of CE-SCTP in the reference scenario depicted in Fig. 6.10. We have assumed a homogeneous demand from users with $\Phi_{\text{MAX}} = 1\text{GB}$ and price sensitivity $\theta = 0.5$. The operator works with a cost per unit $\eta = 0.1\$/MB, which as reported in section 6.3 gives a maximum off-loading ratio $\kappa_{\text{avg}} \approx 0.31$ in order to have profit. In this initial evaluation, every connection is the same; thus a max ratio of 0.31 is employed on each CE-SCTP transfer. We have evaluated three networking architectures: (a) pure ad-hoc wireless, (b) pure cellular and (c) hybrid with four different scheduling algorithms in V2V scenarios. The random

![Figure 6.9: Hybrid Architecture for CE-SCTP](image)

include: (a) RESTful API interface to clients for requesting $\kappa_{\text{avg}}$ threshold, (b) an interface to underlying networks for obtaining network traffic information, (c) HSS/SPR interface for subscriber-specific information that may influence policy decisions.

include: (a) RESTful API interface to clients for requesting $\kappa_{\text{avg}}$ threshold, (b) an interface to underlying networks for obtaining network traffic information, (c) HSS/SPR interface for subscriber-specific information that may influence policy decisions.
scheduler selects one of the available paths randomly, while in RTT and BW schedulers the selection is performed based on minimum RTT or maximum available BW, respectively. Finally, cost scheduler implements a basic CE-SCTP where the selection is based on maximum available BW or minimum RTT, keeping also the corresponding off-loading ratio below the threshold $\kappa_{avg} = 0.31$. We have varied both the number of nodes/connections in the system, and the data rates of those connections.

The most evident result from this evaluation is that multi-homed networks can provide higher throughput, which supports our proposal for hybrid network architecture (Figures 6.11a, 6.12a, 6.13a). However, in low traffic, ad-hoc wireless can support the demand by itself and there is lower need to off-load (Figure 6.11b). This is beneficial for both users and operators. However, as traffic demand increases, ad-hoc wireless network gets congested and schedulers using available bandwidth as indicator start shifting traffic towards the LTE network (Figure 6.12b, 6.13b). The random and RTT-based schedulers show better performance in terms of achievable throughput as the number of nodes/connections increases compared to available bandwidth (BW) scheduler. However they are both non-profitable for the cellular provider in every scenario, which makes them undesirable. The scheduler based only on available bandwidth shows better performance in low traffic demand and its off-loading ratio is well below the $\kappa_{avg}$ threshold. However, when demand increases, the uncontrolled shift of traffic on the LTE network reduced the profit and in certain cases becomes unprofitable. The cost
6.6 Summary

Multihoming is one approach to provide reliable vehicular communications. In this chapter, we take up the challenges involved in supporting multihoming at transport layer. Based on SCTP, we provide an analytical model for dynamic interface selection based on minimum RTT of two paths. Next, using the theorems of economic offloading introduced in [146], we propose a cost efficient SCTP variant. The characteristic of the

scheduler shows similar throughput results as the BW scheduler, but with the added benefit of larger profits for the operator. Further, the distribution of traffic between the two networks among individual users varies depending on the switching scheduler.

While Random scheduler keeps traffic disparity relatively low even for large number of users (Figures 6.14e and 6.14f), RTT and BW schedulers can not control it due to the isolated observations for RTT or estimated BW performed by individual users (Figure 6.14). This is manifested by the large distribution in the box-plots. This issue is important for individual user pricing. The system, from provider’s point of view, might be cost-efficient, i.e. operate below $\kappa_{avg}$, but there are users who are forced to use their LTE allowance more than others, which contradicts with the assumption for homogeneous users.

Figure 6.11: Scenario with data flow 100kbps
Figure 6.12: Scenario with data flow 150kbps

Figure 6.13: Scenario with data flow 200kbps
Figure 6.14: Traffic Distribution among users on the two networks
cost model we employed, is that it considers both the network provider and the end-user perspective and aims to increase the economic benefits of both. The simulation-based analysis has validated the cost-efficiency of the proposed protocol.
Chapter 7

Conclusions and Future Work

7.1 Conclusions

In writing conclusions, the aim is not to reiterate the original contributions, which were covered in Chapter 1. Instead, expose wider reaching implications of the studies that were carried out, with the support of Figure 7.1. As it can be observed, we have investigated or exploited in one way or another, all the levels of the ITS protocol stack presented in Figure 1.4a, apart from security. One of the significant aspects of this work is primarily that it explores all of the layers in ETSI’s protocol stack for ITS systems, contrary to other works, which only address a single aspect. The high level objectives of our work were to address efficiency and reliability issues of inter-networking in vehicular communications, as the title so laconically states. However, the concept of efficiency and reliability are quite abstract and generic. In a bottom-up approach, we demonstrate how these abstract concepts are materialised within our work, and how ETSI’s objectives [16] are achieved.

Starting from the ITS access technologies available for V2X communications, DSRC and LTE are the two pillars upon which the majority of communications are performed. Other types of technologies are used, such as Bluetooth or Radio-frequency identification (RFID); however less frequent as a consequence of low latency and low overhead requirements set by ETSI for vehicular communications. Low latency is interpreted
Chapter 7. Conclusions

Figure 7.1: Visualised summary of contributions

as prioritisation and QoS support, so that emergency messages can reach the destination timely. This is achieved with EDCA mechanism in DSRC and EPS bearers in LTE systems. The communication capabilities of these two technologies are considerably dissimilar. DSRC is based on a decentralised architecture, contrary to the centralised cellular architecture of LTE. In addition, the communication range extends to some hundred meters for DSRC as opposed to several kilometres for LTE. This has an impact on the types of communications and their delay that we investigate through the analytical models presented in chapter 5. We demonstrate that depending on the scenario, the two technologies introduce different amounts of delay. A hybrid system, that can utilise both of them collaboratively can provide lower delays and also reduce overhead. Such a hybrid system we have proposed both for the Location Service and the cost-efficient transport protocol.

Moving up to ITS network and transport layer, both ETSI ITS and IEEE WAVE have proposed a dual stack architecture with GeoNetworking for ITS specific, mainly broadcast, traffic and standard TCP/IP for generic traffic. Initially, our proposal was an enhancement to the GeoNetworking stack with an innovative routing protocol, presented in chapter 3, providing the optimal route towards the destination with minimal
overhead. It incorporates cross-layer information from access layer, as well as facilities, by considering link quality, mobility and utilisation of nodes. ITS Access, Network and Facilities layers are linked through ITS Management layer to provide the required cross-layer information. An Analytic Hierarchy Process (AHP) approach is utilised to combine multiple decision criteria into a single weighting function to make effective forwarding decisions. In addition, AHP analysis contributed to awareness of effects that individual parameters have on the performance of routing. From the results, it is manifested that mobility related information is not the only parameter to be accounted for in forwarding. A cross-layer approach should be considered for efficient performance.

On the other hand, in chapter 6 we investigate generic IP traffic and the limitations of standard TCP/IP stack to support multi-homing. SCTP is one IETF standard that supports multi-homing, though, by default, it is only employed when the primary address becomes unavailable. This increases reliability and robustness of the network against faults, which are two of ETSI’s objectives. There are several extensions to SCTP that exploit multiple network interfaces and paths available to increase throughput and reduce latency. Nevertheless, since the usage of cellular networks is related to higher prices, our proposal for a cost-efficient SCTP alternative is able to tackle the challenge of multi-homing efficiently. By monitoring the traffic flows through DSRC and LTE networks, the mobile provider is able to calculate the off-loading ratio threshold, above which it is operating at a loss. Users are informed for this threshold during the connection set-up, and dynamically switch between the available interfaces. The benefits are both for providers, who are operating with profits, and users who get the most out of their money in terms of performance.

At the ITS facilities layer lay components that provide support to ITS applications and network layer which can share generic functions and data according to their respective functional and operational requirements. In particular in chapters 4 we focus on position and time support component by implementing a centralised Location Service, contrary to ETSI’s approach. ETSI proposes flooding of the network with requests each time a vehicle searches for the position of the destination, which is against its own objectives for low overhead and latency. We advocate that a centralised location server accessible through both DSRC and LTE interfaces can significantly reduce latency and
Finally, at the application level we have investigated other types of efficiency, such as fuel and time for the drivers. Two applications, Green Light Optimal Speed Advisory (GLOSA) and Adaptive Route Change (ARC), were implemented and analysed in parallel with the core work of this PhD [151]. However, since they were outside of the main scope, they are not included in this report. Nevertheless, we mention them here for completeness.

Summarising the work of this thesis, we present briefly how it answers the three research questions stated at the introduction that tackle the objectives of this thesis. The answers will assist future researchers to design efficient and reliable protocols for vehicular systems by providing guidelines on aspects to consider.

**RQ 1:** Which node should be selected as next-hop for efficient and reliable unicast data dissemination? How can we exploit mobility and environment characteristics to improve routing decisions?

**Answer 1:** The proposed cross-layer weighted position-based routing protocol in Chapter 3 with the aid of AHP has provided insights on the impact of mobility and environment characteristics on routing performance. It is shown from the analysis of CLWPR that position information is the most important parameter. However, link quality and node utilisation are also important parameters that improve further the performance of the routing protocol.

**RQ 2:** How to efficiently deliver position information needed for routing? Should the same communication channel be used for data and signalling dissemination?

**Answer 2:** In Chapter 4 we proposed two network architectures for Location Services. The first utilised RSUs equipped with DSRC to carry LS requests/replies to/from the cloud-based location service. The second exploited existing cellular network to LS traffic. The analysis both with simulations and analytical model (Chapter 5) suggested the this split of signalling and data traffic on different communication channels/interfaces is more efficient.
RQ 3: Which interface should be used for cost-efficient data transfers? How could end-users’ and provider’s economic benefits increase with exploitation of multi-homing support?

Answer 3: Several researches have investigated the effect of off-loading cellular traffic to WiFi in order to increase performance in terms of throughput as well as for economic benefits. In Chapter 6, we proposed a cost-efficient transport protocol to dynamically select the network interface that would meet user’s QoS requirements and increase the economic benefits both for users and the network provider.

7.2 Future Work

We have presented and evaluated a variety of techniques to efficiently and reliably support vehicular communications. Although the research has realized its main objectives and answered the research questions stated in the introduction, the challenging nature of vehicular communications requires investigation of other open issues that could extend this work. The research field is still in very active stage and there are several interesting topics remaining for future work.

CLWPR Enhancements

In chapter 3, we have proposed CLWPR and its analysis with AHP. A set of fixed parameters has been produced and used for the comparison with other protocols. However, this approach could be further extended by dynamically selecting the parameters through some sort of inference process such as fuzzy logic, or depending on the situation. For example, delay tolerant applications could require higher delivery ratio without caring about latency, therefore decreasing the importance of utilisation parameter. On the other hand, real-time applications, that require low jitter, could dynamically disable carry’n’forward mechanism or increase its coefficient parameter.

Hybrid routing with LTE and DSRC synergy is another interesting area. We already have demonstrated the benefits of hybrid architectures both for location services and transport protocols. In systems where multi-homing is not supported
by the transport protocol, network could assist by selecting the proper network
interface dynamically.

CE-SCTP Enhancements
An arising topic of great interest with respect to cost effective deployment strate-
gies would also be traffic measurements and models for mobile and nomadic wire-
less data services. During the last years the demand for wide area coverage and
higher data rates have been debated in the research community; now, when traf-
fic demand is increasing, would be an appropriate time for in-depth analyses of
user behaviour. In particular, dependencies between user behaviour and quality
of service would be of great interest to examine.

Further, exploitation of location information of both mobile devices and access
points/base stations, would assist in prediction of connectivity and hence selection
of appropriate network interface.

LTE Direct
One of the advantages of DSRC against current LTE releases is that it supports
broadcast and device-to-device communications, which are required for safety
related applications. With the LTE release 12, LTE-Direct support will become
available which will be a game changer. Alongside the Multimedia Broadcast
Multicast Service (MBMS), they will bridge the gab between DSRC and LTE.
A new delay model should be developed in order to investigate the limits of this
technology and the benefits it can provide to ITS.

Security & Privacy
An important issue in ITS is security and privacy. VANETs are prone to sev-
eral types of security threats from external as well as internal attackers. We
have already mentioned in chapter 2 that the architecture and primitives to cope
with those attacks are in place. However, a hybrid architecture introduces new
challenges to the security and privacy problem. For example, a single key manage-
ment scheme may not be suitable for a hybrid vehicular communication system.
Exploitation of location information in authentication and validation of security
keys could increase their efficiency.
Practicability & Implementability

Furthermore, empirical case studies evaluating the performance of heterogeneous networks are required. Since several important factors e.g., path loss, short term time-dynamics, and other parameters, are difficult to model accurately and fairly across heterogeneous system configurations, empirical case studies and measurements would be needed to know the overall performance of heterogeneous wireless access networks. Last but not least, reasonable practicability and implementability can involve a constraint on what is capable of being done by reference either to practical considerations or to normative considerations.
Appendix A

Supplementary Results

In this appendix we provide supplementary results from the evaluation of CLWPR. The scenario and configuration matches that of Section 3.6 for Scenario #1 (Manhattan grid with synthetic traffic).

A.1 Comparison with topology-based routing protocols

This section presents the comparison of the proposed CLWPR protocol with topology based routing protocols including OLSR, AODV and DSDV that are implemented in NS-3. We evaluate packet delivery ratio (PDR) and average end-to-end delay (E2ED) that relate to the performance of the protocol to provide connectivity in the network. In addition, we measure the simulation time, which relates to the complexity of the protocol. The results presented in Figure A.1a and A.1b show that CLWPR can provide higher packet delivery ratio than every other topology-based routing protocol and lower delay than most of them. OLSR has lower delays than CLWPR as it pro-actively sets up the routes for each node in the network. Further, the implementation of CLWPR exhibits lower elapsed simulation times from all other protocols. This shows the low complexity of the proposed protocol, which only uses local information and performs routing decision per-hop, without the need to calculate end-to-end path.
A.2 Impact of Carry-n-Forward mechanism

In this section we present the impact of carry-n-forward mechanism on the performance of the proposed protocol. Specifically we change the caching time limit from 2sec to 5sec and evaluate its effect. Caching improves by as much as 15% the packet delivery ratio at high mobility scenarios with 5sec caching limit (Figure A.2a). However, the impact on delay is significant as it can be clearly seen in Figure A.2b.
A.2. Impact of Carry-n-Forward mechanism

(a) Packet Delivery Ratio

(b) End-to-End delay

Figure A.2: Performance evaluation of Carry-n-Forward mechanism in Scenario #1
Appendix B

Publications

This list presents published parts of this PhD in journal articles, conference proceedings and technical reports.

   doi: [10.1002/wcm.2578](https://doi.org/10.1002/wcm.2578).


Bibliography


[22] “DRIVE C2X FP7 project,” http://www.drive-c2x.eu/project


[77] H. Menouar, M. Lenardi, and F. Filali, “Movement Prediction-Based Routing...


Propagation Modeling in Urban Environments for Vehicular Ad Hoc Networks,”


