Towards Reliable Geographic Broadcasting in Vehicular Networks

Quincy Tse

A thesis submitted in fulfilment of
the requirements for the degree of
Doctor of Philosophy

Faculty of Engineering and Information Technologies
The University of Sydney

2014

Abstract

Vehicular ad hoc Networks (VANETs) use wireless data communication technologies to allow elements of the road systems to communicate amongst each other, with the aim of improving road safety. In VANETs, vehicles broadcast safety-related messages, including vehicle positions and road conditions, to neighbouring stations. This helps to improve drivers' awareness of the road situation beyond their sensory ranges. VANETs are highly susceptible to channel contention, which can degrade these systems' reliability.

This thesis first addresses the issue of channel use efficiency in multi-hop geographical broadcasts (geocasts) in the VANET environment. Geocasts are used by safety applications to disseminate vehicle status and thus require high reliability. Constant connectivity changes inside a VANET make the more efficient routing algorithms unsuitable. This thesis presents an adaptable, channel-efficient geocast algorithm that uses a metric to estimate the ratio of beneficial to redundant and irrelevant packet reception. Using this metric, relays are selected using a delay-based priority scheme. Through computer simulations, it is demonstrated that this algorithm has more efficient channel use than the farthest-station-first family of strategies, being able to achieve comparable packet reception with lower interference. It is also capable of adapting to channel load by adjusting a single parameter, allowing the algorithm to mimic more efficient algorithms for certain channel situations.

Second, this thesis presents a method of estimating channel load. This technique can be used to provide feedback for moderating the load offered to the

network, including the adjustment of retransmission algorithm parameters. This thesis presents a theoretical model that identifies the relationship between channel load and the idle time between transmissions, which is then exploited to estimate channel contention. Through computer simulations of the IEEE 802.11 Distributed Coordination Function (DCF), channel estimators implementing Bayesian inference and configured with the observation probabilities derived from this model show higher accuracy and much faster convergence time to the steady state than an existing method presented by Bianchi and Tinnirello that observes packet collisions.

Furthermore, the effects of unsaturated stations on the performance of this estimator are investigated. This thesis demonstrates that unsaturated stations have small but observable effects on the relationship between channel contention and both idle slot counts and collision probabilities. Through simulations, it can be shown that estimators observing idle slot counts to determine an "equivalent" number of concurrent saturated stations are less affected by unsaturated stations than observing just channel busy probabilities. An extension to the DCF Markov model, accounting for unsaturated stations, is also presented but this model is shown to be not viable due to its increased complexity and floating point errors.

Third, this thesis improves the reliability of VANETs in multi-hop geocasts by adapting to instantaneous channel conditions. A modified version of the channel estimator is presented, enabling it to track instantaneous channel conditions. Coupling this tracking estimator to the geocast algorithm produces a closed-loop load-reactive geocast system. Through computer simulations, this closed-loop system is not only shown to be more efficient in channel use, but is also able to automatically adapt to channel contention. This system is observed to self-correct suboptimal retransmission decisions as well.

Finally, this thesis identifies the non standard compliant behaviours of a number of commonly used network simulators when simulating broadcasts. Given broadcasts are the dominant form of safety message dissemination in VANETs,

such misbehaviours can impact on the validity of VANET research, and must therefore be accounted for. This thesis first demonstrates that these errors exist in all the simulators tested, and subsequently describes in depth the error in ns-3, culminating in a set of workarounds that allows simulation outcomes to be interpreted correctly. These workarounds are applicable to most versions of the ns-3 simulator.

Acknowledgements

To my family

For providing me with the opportunity to pursue my dreams. Without their support, I would not have been able to start, let alone complete, this thesis.

To my lovely wife

For encouraging and supporting me in the completion of my thesis. Without her, this thesis would not have been completed. I am extremely grateful for her understanding and support, especially throughout the final stages of thesis writing.

To my supervisors

First and foremost, to Professor Björn Landfeldt, my primary supervisor for 4 years. Thank you also to Professor Albert Zomaya, who supervised the final stages of my candidature.

To Dr Javid Taheri and Dr Weisheng Si — thank you for stepping in and providing much needed guidance, helping me with my thesis at the final stage, and doing so on top of their normal workload.

Thank you also to Dr Zainab Zaidi, who had provided me with clear directions when all my supervisors were away, and helped me on my first publication resulting from this work.

To my labmates

Saeed Bastani and Emma Fitzgerald from the School of IT for providing me with intellectual discussions in VANETs when the vast majority of our research group think in the cloud; and Yan Shvartzshnaider, Mentari Djatmiko and everyone in NICTA for giving me a chance to mingle socially and expand my research interests... and for organising pizzas. Thank you all who have helped proofreading my thesis.

To NICTA

The work in this thesis was partially funded from a scholarship provided by NICTA. NICTA also provided me with opportunities to broaden my knowledge into topics outside my narrow research focus. NICTA is funded by the Australian Government as represented by the Department of Communications and the Australian Research Council through the ICT Centre of Excellence program.

Contents

List of Tables

List of Figures

List of Algorithms

Chapter 1

Introduction

In a society where people regularly travel using various forms of road transport, traffic accidents and congestion are extremely costly in terms of productivity loss, energy consumption and travellers' health. As the number of vehicles using public roads increases, the impact of traffic incidents also increases.

Vehicular ad hoc Networks (VANETs) have been proposed as the next-generation road safety system, allowing vehicles and drivers to extend their awareness of the road situation, thereby improving the reaction time when unexpected events occur. A VANET acts as a large collection of distributed sensors connected by a wireless network, with each vehicle sharing relevant information to all others. The success of these systems depends heavily on the reliability of the underlying wireless communication network — low reliability can affect the accuracy of the safety systems. Too many false positives generated by such systems may lead to high cognitive load and general inconvenience to the driver; false negatives can fail to prevent hazardous situations, leading to mistrust of these systems.

In wireless networks, the typical causes of low reliability are physical-layer effects such as shadowing and fast fading, and MAC-layer packet collisions caused by hidden terminals and channel contention. Using MAC-layer techniques, this thesis aims to improve the overall network reliability in high-load environments affected by shadowing by using cooperative retransmissions. It also highlights non-standard behaviours observed from commonly used simulator packages for

broadcast mode messages, with a set of workarounds proposed for using and interpreting outcomes from ns-3 simulations despite the non-standard behaviours in the simulator.

1.1 Outline

This thesis aims to present a closed-loop load-reactive geocast system. The individual components of this system are themselves novel, and this thesis first presents, evaluates and discusses these separate parts before presenting the overall design of the entire system.

Following this introductory chapter, this thesis first provides a background of technological developments in the road transport and vehicular technologies in Chapter 2. This chapter outlines the key evidence supporting the need for these networks, and reviews the technology which may be used to implement them. An in-depth explanation of IEEE 802.11 broadcast procedures, which is relevant to the theoretical analysis in later chapters, is also included.

A review of the current and previous research into VANETs is then provided in Chapter 3. This chapter discusses the works related to the three distinct areas covered by this thesis — techniques to improve packet reception in VANETs (including routing algorithms), methods to gauge and adapt to channel contention, and evaluations of simulation packages.

Chapters 4 to 8 present the works conducted for this thesis. Chapter 4 presents and evaluates a static, interference-aware, distributed geographic broadcast algorithm. Chapters 5 and 6 deals with estimating channel contention in networks consisting of saturated and unsaturated stations respectively. Chapter 7 then links the work in these previous chapters into a closed-loop load-reactive geocast system.

Chapter 4 addresses the issue of packet reception ratio in DSRC-based VANETs being below the requirement set by the U.S. Department of Transport. A retransmission metric and a retransmission algorithm using that metric are presented,

including a variant that also adapts a station's transmission range as needed. These algorithms were tested against other similar algorithms in computer simulations, showing their flexibility in adapting to channel condition as long as the algorithm parameter can be adjusted dynamically.

The need to dynamically adjust the algorithm parameter leads to the channel contention estimation technique that is presented in Chapter 5. In this chapter, the technique of estimating channel contention by observing the length of the idle period between transmissions is presented. It is intended that, by sensing the current channel contention level, a station can better moderate the load it offers to the network, for example by adjusting the retransmission algorithm parameter. As part of this technique, a Markov model of a DCF broadcasting station is presented and used to discover the relationship between channel contention and observed channel statistics. Computer simulations demonstrating the accuracy of the DCF model and the viability of the contention estimation technique are presented.

The investigation in Chapter 5 focuses on saturated stations, which are uncommon in real VANETs. Chapter 6 investigates the effects of unsaturated stations on the network observation. To compare these observations, a measure called "Equivalent Saturated Node" (ESN) is defined to both describe the level of saturation of a station and the level of contention in a network. Through computer simulations, it is shown that unsaturated stations cause observable differences amongst networks with the same level of contention. It is however also demonstrated that such differences do not appear to greatly affect the channel contention estimates from the estimation techniques presented in Chapter 5. Finally, an extension to the DCF model accounting for unsaturated stations is presented, but is found not to be viable due to its complexity and floating point errors.

Having presented and analysed both the techniques to improve packet retransmission decisions and the mean to determine channel contention, Chapter 7 couples these components together to form a load-reactive geocast system. This

chapter first confirms the validity of using ESN to predict packet reception before presenting an algorithm that couples the load sensing component to the interference-aware geocast algorithm. This investigation clearly shows that the interference-aware geocast algorithm is efficient in using the radio channel. In addition, an emergent behaviour in the closed-loop system such that the geocast system would self-correct suboptimal retransmission parameters calculated by the load adaptation algorithm is observed.

As the work in this thesis relies heavily on computer simulation of broadcast packets, the accuracy of these computer simulations is important. Chapter 8 validates simulation packages by comparing their results of broadcast mode transmissions to the expected theoretical values. It is demonstrated that the commonly used simulator packages all showed non standard-compliant behaviours. Further investigations in the chapter identify caveats that allow results from the ns-3 simulator to be correctly interpreted.

Finally, this thesis is concluded in Chapter 9, with suggestions on future works arising from this thesis identified.

1.2 Contributions

- **I developed a metric that ranks wireless stations for relay preferences.** This metric considers both the extra coverage a station provides, and the interference it introduces by retransmitting. In addition, this metric is computed from the potential relay without the need for coordination amongst other stations. It is independent of the actual retransmission algorithm and hence can be used by other algorithms for prioritising relays.

- **I implemented and evaluated a retransmission algorithm utilising the retransmission metric.** The metric was implemented in an ns-3 simulation, together with a delay-based relay selection algorithm. Simulation results showed that the metric is capable of selecting good relay stations

in order to cope with high network contention scenarios, with the scalability controlled by a dynamically adjusted parameter.

- **I investigated the relationship between wireless channel contention and observed MAC-layer idle slot counts.** A Markov model of the MAC-broadcast DCF was constructed in this investigation. Numeric solutions to the model provides a mapping between the probability distribution of interframe idle slot counts and the channel contention in terms of the number of concurrent saturated stations. This mapping can be used by the MAC layer to estimate channel contention, in order to adjust MAC parameters and/or to provide feedback to upper layers for moderating the offered load onto the network.

- **I demonstrated and evaluated a passive technique for estimating channel contention using simple Bayesian inference.** Using the probability distribution computed from the Markov model, the technique of estimating contention through observing idle slots was compared to Bianchi *et al.*'s MAC-level contention measurement technique using computer simulations. I have shown that estimates from this technique converge to the scenario parameter quicker and is more accurate.

- **I demonstrated the effects unsaturated stations have on the relationship between the wireless channel contention and the observed idle slot counts and their impacts on channel contention estimation techniques.** Here, channel contention is defined as the sum of individual stations' saturation across all stations in the network. Through simulation, I showed that station saturation has a small but observable effect on both the distribution of idle slot count and the collision probability. There are minor impacts on the estimators' channel contention estimation accuracy as well as slight lengthening of time before the estimates stabilise. I have also shown that the technique of observing idle slot counts is more

resilient to errors caused by unsaturated stations.

- **I have validated the usefulness of the ESN metric in predicting packet non-reception.** Statistical analysis on computer simulation results showed that a simple threshold test on observed ESN value has a very high Negative Predictive Value. This means that ESN can very accurately predict packet non-reception.

- **I have used statistical techniques to provide further evidence on the efficiency of the interference-aware geocast algorithm.** Statistics on the ESN-based threshold test shows that the test has a higher Positive Predictive Value on the interference-aware geocast algorithm than a greedy distance-based technique. This provides further evidence that the geocast algorithm is more efficient in using the channel to improve packet reception.

- **I have designed and evaluated a geocast algorithm that changes its behaviour in reaction to channel contention.** This algorithm uses outputs from the passive idle slot-based channel estimator to determine whether rebroadcasts should be increased or suppressed, and adjusts the retransmission parameter of the interference-aware geocasting algorithm automatically. This allows the algorithm to adapt to channel conditions without the need for manual intervention.

- **I have identified high discrepancies between outputs of different commonly-used network simulator packages.** The discrepancies amongst the simulators are likely to be caused by errors in the implementation of the IEEE 802.11 MAC-layer broadcast behaviour.

- **I have evaluated the impact of ns-3 broadcast-mode misbehaviour.** By comparing the simulation results to theoretical predictions, I have shown that the misbehaviour observed from ns-3 simulations of broadcast mode

IEEE 802.11 transmissions has a small impact in terms of application-layer performance, but has major effects on algorithms that rely on MAC-layer observations such as collision probabilities and idle slot counts. The observations and analysis is applicable to all versions of ns-3 at least from ns-3.4 to ns-3.15. (It is most probable that the workarounds are also applicable to releases after ns-3.0.4 when the YANS Wifi model [1] is first introduced.)

1.3 Publications

- **Tse, Quincy**, "Improving Message Reception in VANETs," in Proceedings of *Mobile Systems PhD Forum, 2009 International Conference on*, Krakow, Poland, Jun 2009.

- **Tse, Quincy** and Landfeldt, Björn, "Interference-Aware Geocasting for VANET," in Proceedings of *World of Wireless, Mobile and Multimedia Networks, 2012 IEEE International Symposium on*, pp. 1-6, San Francisco, CA, USA:IEEE, 25-28 June 2012.

- **Tse, Quincy**, Si, Weisheng and Taheri, Javid, "Estimating Contention of IEEE 802.11 Broadcasts Based on Inter-Frame Idle Slots," in Proceedings of *Local Computer Networks Workshops (LCN Workshops), 2013 IEEE 9th Conference on*, pp. 120-127, Sydney, Australia:IEEE, 21-24 October 2013.

Overview

This chapter provides background information on ITS and related fields, leading to the development of VANETs. In addition, an introduction to the IEEE 802.11 standard relevant to this thesis is presented.

Chapter 2

Background

Road transport now forms an integral part of many people's lives such that any disruptions to the road system can be extremely costly. The Australian Bureau of Transport and Regional Economics projected that the avoidable social costs of traffic congestion in Australian capital cities will rise from $9.4 billion in 2005, to $20.4 billion in 2020 [2]. The U.S. Government estimated traffic congestion to have costed US$87.2 billion in 2010 [3]. In addition, traffic accidents amounted to a loss of $17 billion in 2003 alone, resulting from material costs, social costs and productivity losses [4]. In 2005, 1,627 road fatalities and a further 31,204 hospitalisations for an average of 4.4 days were recorded in Australia, representing 8.0 deaths and 153 hospital admissions per 100,000 population [5,6]. These figures are very close to the OECD median. It is therefore both socially and economically important to improve road safety and efficiency.

Research into technological improvements in both road traffic management and vehicular safety started in the 1960s. Outcomes from these research had greatly enhanced both the efficiency of road transportation and reduced accident rates by using on-road sensors and better management algorithms, and had improved vehicle safety through the development of in-vehicle safety devices.

2.1 Technologies in traffic management

The efficient operation of the road transport network is extremely important in any city. In addition to avoidable social costs, traffic congestion also has impacts on the economy in terms of productivity loss, and are associated with increased traffic accident rates. Weisbrod *et al.* estimated an increase of around US$250 million in the cost of business in the Chicago CBD alone if traffic congestion caused a 6% delays in truck delivery [7]. Traffic congestion is also a major contributor to pollution in terms of emissions of greenhouse gases, particulates, nitrites and other pollutants. For these reasons, improvements in traffic efficiency is a priority for governments across the world.

One of the earliest technological developments in traffic management is the traffic light. Since their introduction, the operation of traffic lights has increased in complexity, improving the overall network-wide traffic efficiency. The traffic lights have improved from using simple timer-based scheduling to sensor-triggered phase changes. Today's traffic lights are not only sensor actuated in order to adapt to varying traffic levels, they are also interconnected and are coordinated by central systems such as SCAT [8] and SCOOT [9], improving traffic efficiency in terms of the expected number of red lights encountered, trip times and/or overall network speed and throughput. The current generation of traffic management systems take input from a wide range of sensors, including the inductive loop detectors, magnetometers, pneumatic tube sensors and video cameras, and control traffic using traffic lights, tidal-flow control systems, variable message boards, variable speed limit signs, as well as TV and radio broadcasts and telemetry systems [10, 11].

These current generation systems are unfortunately limited by the specificity of control messages, and their inability to obtain the travel intentions of the individual vehicles. Current traffic control devices affect all vehicles in the target area, lacking fine-grain control. Furthermore, travel intentions of vehicles can only be gauged from aggregated coarse-grain information sources such as population

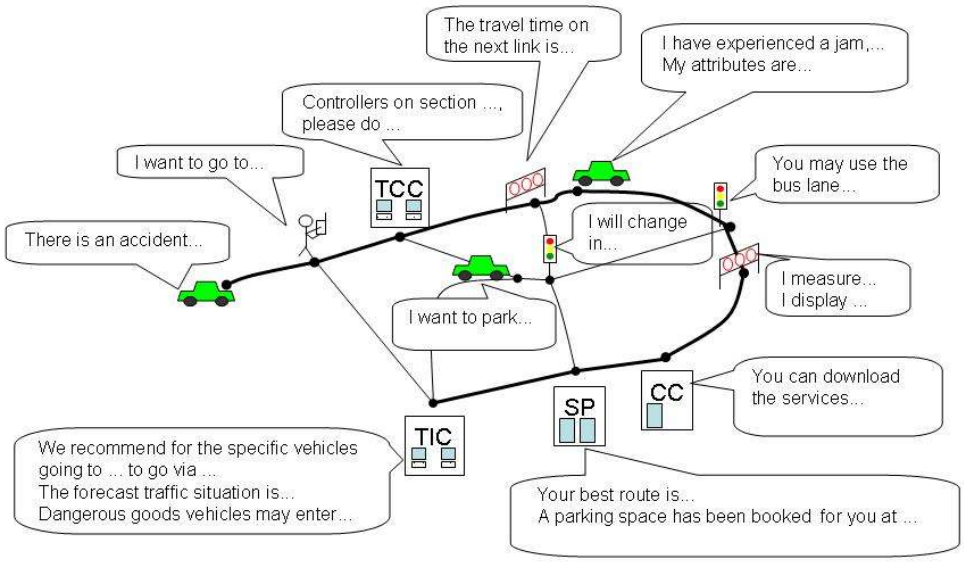


Figure 2.1: NICTA's vision of intelligent transportation systems in year 2020, where the road infrastructure is able to send to road users up-to-date and relevant information, and road users can report relevant information back to traffic authorities [12].

census data, as well as traffic flow, speed and demand data from the sensors deployed. These sources are poor at predicting traffic demands in presence of extraordinary situations such as congestion and accidents. The collection and processing of trip origin-destination data can greatly assist in predicting the traffic demand, improving traffic scheduling.

The next generation of ITS systems aim to improve the granularity, detail and accuracy of both sensor inputs and traffic control capabilities. These capabilities can only be achieved if bidirectional communication can be reliably established between individual vehicles and traffic control devices. Figure 2.1 shows a typical vision for ITS in year 2020. These capabilities can be enabled by technologies such as wireless networks.

2.2 In-vehicle safety devices

In addition to enhancements in the road infrastructure and network control systems, improvements to the safety of individual vehicles are also critical for road transportation. The vehicular technology community had introduced a number of in-vehicle safety devices since its inception, leading to improved safety outcomes. Worldwide statistics have shown reduction in motor vehicle fatalities correlating to the introduction of these safety technologies (Figure 2.2). These safety devices can be broadly divided into passive and active safety devices.

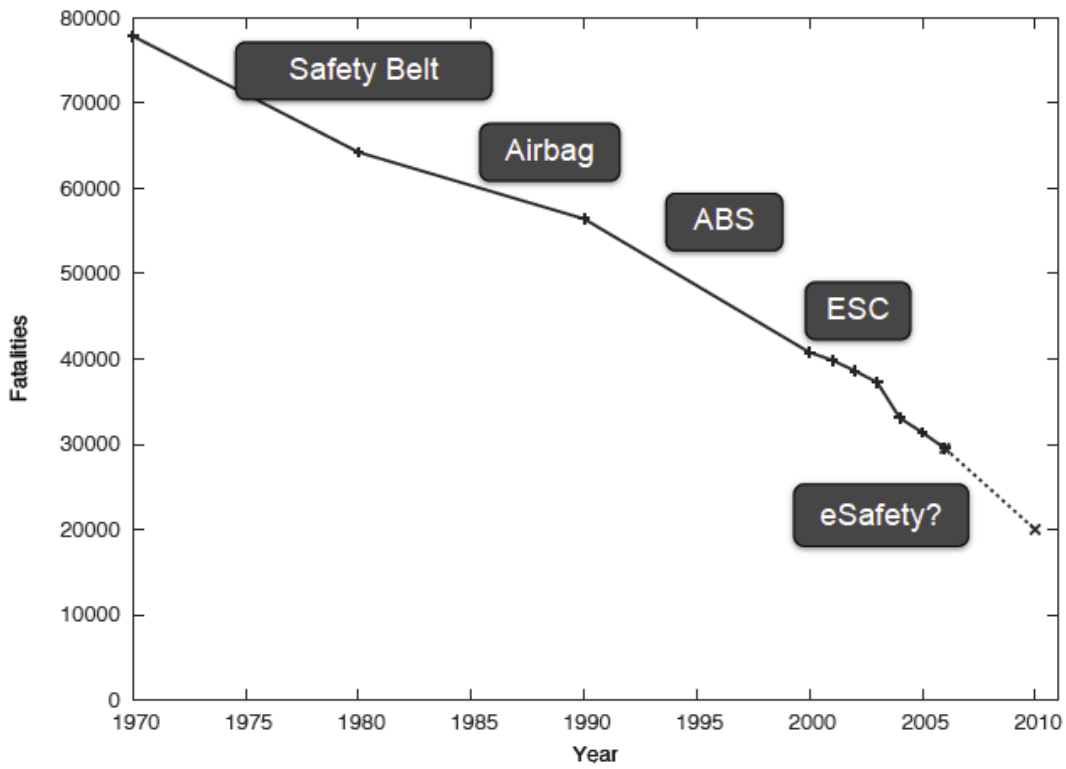


Figure 2.2: European (E15) road fatality statistics based on ERF Road Statistics 2008, showing the approximate periods of safety devices introduction, with projected fatality reduction for 2010 as a result of "eSafety" technologies [13].

Passive devices are designed to reduce the severity of the consequences should accidents occur. These devices include seat belts, pretensioners and airbags. Seat belts are designed to restrain the occupants, reducing the likelihood of the occupant being ejected from the vehicle during collision or from hitting parts of the vehicle. Pretensioners are designed to minimise the amount of movement

permitted by the seat belts, thereby improving the protection offered by seat belts. While the correct wearing of seat belts have been shown to save lives during traffic accidents, the incorrect wearing of it is known to cause severe abdominal injuries in relatively minor traffic accidents [14].

Similar to seat belts, airbags are passive devices designed to reduce the consequences of collisions. SRS and SRP airbags are designed to deploy when they are triggered during a collision, creating a "pillow" of air to reduce the impact to the occupants' bodies against the vehicle components and/or ground. Airbags also require occupants to be properly restrained and positioned. Inappropriate positioning and/or restraint of occupants (*e.g.* young children in front passenger seat) have been known to cause death when airbags are activated [14].

Unfortunately, these passive devices can only reduce the severity of the injuries caused by accidents, and cannot prevent the occurrence of accidents. For this reason, much research and development is now focused on the prevention of accidents using active safety devices such as electronic stability control.

Active safety technologies continuously monitor the driving situation, and initiate behaviours (such as warnings or direct control of the vehicle) to reduce the likelihood of accidents. Well-known examples of this category include electronic stability control, anti-lock brakes and traction control. These features aim to improve vehicle traction, thus controllability of the vehicle, during turns, braking and acceleration respectively, by autonomously moderating engine output and brake pressure. These systems have been assessed as among the most important safety features that governments should take action on its adoption [15]. Unfortunately, these technologies can only react to the vehicle's own condition — they are unable to react to its surroundings.

More recently, various Advanced Driver Assistance Systems (ADAS) have been developed, enhancing a vehicle's awareness of its own immediate surroundings. Forward collision avoidance technology is one such system which had already been commercialised. This technology uses various longer-range sensors such as

radar, LIDAR, infrared, ultrasound and/or computer vision to allow vehicles to detect the distance they are from the objects in the front, and can be used to prevent or reduce the severity of frontal collisions. A government funded study using computer simulation of reconstructed crash scenes had identified such systems as being capable of reducing overall fatal collisions by 30% and injuries by 40% [16]. This, and similar technologies such as adaptive cruise control, increase the vehicle's awareness from just itself to also include its immediate surroundings, thereby allow better preparedness for potential hazards. These technologies are still unable to react to situations that are more than a few meters away from the vehicle.

The next generation of vehicular safety systems not only rely on sensors mounted on individual vehicles, but also integrate inputs from sensors in other vehicles and in the environment. This allows vehicles and drivers to extend their awareness of the road situation beyond the range of any single vehicle sensor and beyond human capabilities, potentially greatly enhancing the accident prevention capabilities of the vehicle. In order to use sensor data from outside the vehicle, reliable wireless communication amongst road elements is essential.

2.3 Vehicular ad hoc networks (VANETs)

Recent developments in wireless computer networking technologies had provided the opportunities for vehicles to become connected to each other as well as to the road infrastructure, providing the technological platform for the next generation traffic management and vehicle safety systems. The wireless network on which vehicles and road infrastructure communicates is termed Vehicular ad hoc network (VANET).

VANETs consist of mobile stations placed in vehicles (OBE, "On Board Equipment") and fixed stations typically collocated with road infrastructure (RSE, "Road-Side Equipment"), wirelessly connected in an opportunistic manner. These networks are used mainly to facilitate autonomous communication amongst ve-

hicles and between vehicles and road infrastructure. Similar to mobile ad hoc networks (MANETs), VANETs contain stations that move over time, but differs from MANETs in terms of mobility pattern. VANET stations typically follow a more rigid movement pattern (*i.e.* along roads) at high velocities. The high station velocity reduces the amount of time any pair of stations remain connected — Jerbi *et al.* measured the average time a pair of stations remained connected to be approximately 60 seconds [17] — making some MANET routing algorithms unsuitable for VANETs.

The sharing and the possible distributed processing of information amongst vehicles also makes VANETs similar to wireless sensor networks (WSNs), with the added bonus that battery power is not a constraint on VANETs. This difference makes the power-saving features in most WSNs unnecessary, with the trade off inherent in most WSN routing protocols negatively impacting on network performance. Furthermore, the mobility pattern in WSN are different to MANET and VANETs, with WSN stations typically not moving much in space, and connectivity changes are typically due to power-saving features. For these reasons existing broadcasting techniques developed for these networks may also not be applicable for VANETs.

In terms of the applications of VANET, the main purpose of systems utilising VANETs is to improve the safety and efficiency of the road system, by giving drivers and vehicles better situational awareness, as well as finer-grained traffic interventions. A secondary objective of VANETs is to facilitate so called "comfort applications" — applications that are not safety critical and do not improve traffic, but are useful and/or convenient to drivers or the greater population so as to help promote the uptake of the technology. Table 2.1 outlines some of the proposed applications that can use VANET as the underlying communication technology.

VANET applications typically have two modes of operation — vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I). V2V applications are those

Table 2.1: Examples of proposed VANET applications

Safety	Cooperative Collision Avoidance
	Post-Collision Warning
	Wireless Traffic Signage
	Wireless Traffic Lights
	Emergency Vehicle Warning
	Emergency Vehicle Signal Preemption
Efficiency	Adaptive Cruise Control
	Cooperative Platooning
	Accident Reporting
	Origin-Destination Reporting
	Electronic Toll payment
	Dangerous Goods Tracking
Comfort	Repair Notifications
	Wireless Diagnostics
	Software Updates
	Cooperative Headlight Aiming
	Parking Spot Reservation
	Enhanced Route Guidance/Navigation
	GPS correction and position improvement
	Internet Services
	Mobile Media

where relevant information are shared amongst vehicles via wireless links. Typical uses of V2V are the sharing of vehicles' positions and the transmission of emergency warning signals. V2I applications are those where information are communicated between vehicles and road side equipment, which may or may not forward the information to other entities. Example of V2I include wireless traffic lights where signal phase data are communicated from the traffic lights themselves directly to the vehicles, and parking spot reservation systems where parking intentions are communicated by the vehicles to the car park operators via an RSE.

2.3.1 Safety applications

Safety is the primary focus of VANET technologies. The intention of VANET-based safety applications is to provide drivers and vehicles an extended hazard detection distance by sharing relevant information, allowing vehicles and drivers more time to react to the hazards. It was shown that approximately 80% of

vehicle collisions happened at intersections or in low-visibility areas, and 70% of the accidents were caused by the failure to recognise hazards in time [18]. Solutions that can improve the hazard perception time could therefore potentially improve the safety of the road system. The U.S. National Highway Traffic Safety Administration and the Crash Avoidance Metric Partnership (CAMP) identified a few priority applications and their communication requirements (Table 2.2) [19], some of the applications identified had been combined in the discussions in this chapter. The majority of the safety applications proposed rely on geographically-bound broadcast (geocast) messages conveying relevant information. The use of geocast allows much more efficient message transmissions that apply to all vehicles in a given area.

In the V2I settings, applications such as wireless traffic signage and wireless traffic lights have been proposed. The idea of wireless traffic signage is similar to that of the traditional static or variable traffic signs, except the information is disseminated as data packets directly to each vehicle. This allows the vehicles to take relevant actions as necessary, for example, adjusting traction control parameters when a warning messages from a wireless traffic sign indicating slippery conditions on the road is received. One interesting extension of this concept is to transmit control signals that are machine-readable, but may be too complex for humans. Fitzgerald *et al.* presented a concept whereby road segments are not limited by vehicle speed, but rather by "risk" calculated from various physical factors such as driver age, vehicle condition, blood alcohol concentration, *etc.* [20]. Such "risk limit" would be extremely difficult for humans to interpret on the fly, but on-board computers could easily convert the limit obtained from the wireless road sign into a more human-friendly form such as speed limit.

The concept of wireless traffic lights is similar to wireless traffic signs. There had been proposals suggesting that traffic lights, being located in a good position to connect to most vehicles on the intersection, be used to aggregate traffic information, and to disseminate summaries to all vehicles in that intersection for

Table 2.2: Priority VANET safety applications in USA and their communication requirements [19]

Application	**Comm. Type**	**Rate**	**Max. Latency**	**Range**	**Data Transmitted**
Traffic Signal Violation	V2I	10 Hz	100 ms	250 m	signal phase, timing, position, direction, road geometry
Curve Speed Warning	V2I	1 Hz	1000 ms	200 m	curve location, curvature, slope, speed limit, surface
Emergency brake lights	V2V	10 Hz	100 ms	200 m	position, heading, velocity, acceleration
Pre-crash sensing	V2V	50 Hz	20 ms	50 m	vehicle type, position, heading, velocity, acceleration, yaw rate
Forward collision	V2V	10 Hz	100 ms	150 m	vehicle type, position, heading, velocity, acceleration, yaw rate
Left (right) turn assist	V2X	10 Hz	100 ms	300 m	signal phase, timing, position, direction, road geometry
Lane-change warning	V2V	10 Hz	100 ms	150 m	position, heading, velocity, acceleration, turn signal status
Stop sign assist	V2X	10 Hz	100 ms	300 m	position, velocity, heading

safety purposes. (For example identifying and disseminating warning message about high speed oncoming vehicle likely to run red light.) Other useful information that can be included are signal phase information that allows individual vehicle to decide whether it should attempt to cross during the current green phase, as well as coordinating vehicle platoons to facilitate more efficient traffic. A more radical vision of the wireless traffic lights sees that vehicles would independently decide whether to cross an intersection without explicit red/green phases, with the decision making facilitated by the wireless traffic lights [21]. Such a system would greatly improve the traffic throughput of the intersection.

A third V2I application proposed is emergency vehicle signal preemption. This refers to the ability for emergency vehicles to request traffic light phase change in order to enable the emergency vehicle to not need to slow down at red lights. It can also reduce the risk of oncoming traffic that is currently being shown the green light not knowing about the approaching emergency vehicle and hence not give way. This third application, while useful, can already be implemented using current technologies: since many emergency services already track their fleet using various positioning techniques, one only needs to install a gateway between the emergency service control centre and the traffic management centre in order to request traffic light phase change.

Most of the networking aspects in VANETs tend to focus on V2V scenario. The common examples cited in the literature includes various variants of Cooperative Collision Avoidance (CCA) [22, 23] and post-collision warning (also called Cooperative Collision Warning (CCW)) [24,25]. A third commonly cited example is the Emergency Vehicle Warning.

Cooperative Collision Avoidance aims to prevent collisions by enabling the sharing of vehicles' knowledge of road situation (or at a minimum, knowledge of themselves) amongst each other. Even though more information can be sent, it is assumed that, at a minimum, beacon packets containing the vehicles' positions, directions and velocities are broadcasted every 100 ms. By collecting

these beacons from its neighbours, a vehicle can derive a map of its surroundings to determine and react to hazards. Additional information that may also be broadcasted include the status of vehicles and their sensors (*e.g.* whether their turn indicators are on, whether surface traction is poor at a road segment). Depending on the way CCA information is used, it can be further classified into more specific types such as left/right turn assistance, merge assistance, blind spot warning, extended brake lights, *etc.* Timing requirements for such applications had been investigated in works such as [26]. In terms of network requirements, CCA can typically withstand occasional packet losses due to the inherent redundancy in consecutive position data, but these packets are delay sensitive. El Batt *et al.*have identified that consecutive packets can be lost not infrequently even in highway environments, and as vehicle density increases, single-hop reception ratio deteriorates quickly [23]. This points to the need for further research into improving the communication system reliability.

Unlike Cooperative Collision Avoidance systems, which actively tries to prevent collisions from occurring, post-collision warning (also called Cooperative Collision Warning or CCW) systems attempt to prevent secondary collisions. Typical example used to explain and investigate such system involves the scenario where a vehicle had suddenly deployed its airbags (indicative of a collision) — the vehicle involved in the accident would generate a relevant warning and broadcast it to its neighbours. This allows all its neighbours to react to the changed situation, *e.g.* prepares oncoming vehicles for collision because it is too close to avoid chain collisions, activate brakes for vehicles close enough to be threatened but have time to react, or to issue a warning to the following drivers through human-computer interfaces. CCW messages are not only time critical, but also requires high reception reliability. Biswas *et al.* have analysed the requirements for such systems, and showed that high packet reception ratio is important in the operation of such systems [24]. (Note: Biswas *et al.* referred to these system as "CCA" in their paper.)

Emergency Vehicle Warning systems have also been proposed as an extension to the existing lights-and-siren systems for warning road users of oncoming emergency vehicles. These systems extend the existing systems by broadcasting digital warning messages to each vehicle in range, which can in turn inform their drivers. This system helps mitigate the problems faced by the newer model vehicles that have better sound isolation, which, when combined with loud music being played in the vehicle, would mask the siren sounds from the emergency vehicle. These systems would have similar data requirements as with CCA systems, except the required transmission area may need to be longer and be skewed towards the front of the emergency vehicle. The shaping of antenna beams in such situation is a physical-layer consideration, and is beyond the scope of this thesis.

All the aforementioned safety applications require very high reliability in terms of delays and/or reception ratio. Poor network reliability may cause hazardous conditions to remain undetected and/or safety messages not received and processed until too late. This thesis aims to improve the overall reliability of VANETs in the V2V setting. While the current standards for VANETs assume CCA beacons are only transmitted to single-hop neighbours, due to radio propagation difficulties, multi-hop forwarding of messages may be advisable for some safety applications. Through simulations, Chen *et al.* [27] showed that one-hop packet reception falls quickly below the requirements specified by the US Department of Transport. This shortfall is also demonstrated in proof-of-concept prototype studies in the U.S. PATH project [28]. The work in this thesis tries to improve packet reception for CCA and CCW by using other vehicles to forward packets, and tries to reduce the channel load caused by these retransmissions so as not to overwhelm the channel.

2.3.2 Traffic management applications

Another major focus of VANET applications is to improve traffic throughput. V2V applications in this area include technologies such as Adaptive Cruise Con-

trol and Cooperative Platooning. These features use VANETs to communicate relevant information between vehicles, increasing traffic efficiency by enabling tighter spacing between vehicles [24]. Both the US PATH project and Honda have separately demonstrated tight platooning of vehicles with headway of 4–6.5 m using a combination of in-car sensors and V2V communication [29]. Such tight headway can greatly enhance the throughput of the road traffic network.

The more useful traffic management applications tend to be V2I. Data communication enabled by VANETs allows more detailed information such as detailed accident reports, origin-destination information, *etc.* to be gathered by traffic authorities, allowing better dispatching and scheduling decisions to be made. Notwithstanding the current technologies already implemented, applications such as electronic tolling and dangerous goods tracking were also envisaged for this system. While this thesis does not address the specific concerns related to the communication requirements for these applications, the retransmission and channel load detection techniques presented may be applicable to these applications.

2.3.3 Comfort applications

VANETs are also designed to support a range of other non-safety critical and not management-related applications, broadly referred to as "comfort applications". These applications may be the key drivers for the adoption of VANET technologies, allowing these technologies to gain sufficient market penetration in order for the safety and management applications to operate. Applications envisaged include vehicle maintenance (repair/service notifications, diagnostics and updates), enhanced route guidance and navigation (improved positioning data [30,31], traffic situation-aware routing), parking assistance (parking reservations and directions), as well as general network services including web browsing. However, it must be stated that VANET may not be the most appropriate means to provide certain applications (*e.g.* mobile data services may be more cost-effective for general network services). This thesis does not address issues related to com-

fort applications, which may be extremely variable depending on the individual service requirements.

2.3.4 Other considerations for VANETs

Having discussed the various applications intended for VANETs and their networking requirements, it is also important to be aware of requirements that are not network communication related. Not only do these factors potentially increase the channel load for their implementation, they may also impact on the appropriateness of the underlying communication protocols. Due to the sensitivity of messages passed over VANETs, security and privacy are two major considerations for VANETs. Other factors for VANETs includes data processing requirements, costing and deployment, as well as legislative framework. The impacts from these other factors on communications protocols are not high, and are not discussed in this thesis.

Since VANETs are proposed for safety-related applications, it is extremely important to ensure that integrity, availability and authenticity of messages are safeguarded. Current proposals for the provision of security functions include the use of dedicated hardware and the use of special cryptographic functions, with sensitive keys, *etc.* potentially be transmitted ("updated") over the air. If encryption keys are updated over the air, then these essential communication (and associated protocol overhead) will contribute to the network load and potentially reduces the available resources for safety messages, making the efficient forwarding of safety messages important.

Privacy is also a major concern for VANETs. It is not difficult to track vehicles and infer personal information by collecting CCA messages unless other privacy safeguards are in place. Many privacy-preserving techniques have been proposed, and many involve the use of frequently changing pseudonyms. The use of pseudonyms greatly increases the complexity for CCA applications as these pseudonyms inflate the number of stations seen by pseudonym-unaware algo-

rithms. For example, algorithms that estimate channel load by counting the number of different network identifiers in the local area is necessarily inaccurate if each station may have multiple pseudonyms. The channel load measurement technique presented in this thesis is resilient to pseudonym changes as it relies on the properties of the MAC layer without the need for identifier tracking.

2.4 Technical developments in VANETs

The basic architecture and the protocols used for VANETs have been standardised separately in Europe and in the USA. They are partly compatible architectures, and are currently in the process of harmonisation. Both architectures define a stack of interrelated protocols that form the fundamental communication links, with the European (ESTI) architecture also including a more specific definition of the higher-layer functionality.

The USA commenced investigating vehicular communications in November 2003, with the U.S. Department of Transportation (USDOT) announcing a new initiative called "Vehicle Infrastructure Integration (VII)". The project attempted to specify, design, build and test a small-scale prototype of the proposed system, projected to begin deployment around 2010. The system was evaluated against three sets of application-goals (safety, mobility and "private services") and three sets of system requirements (security, maintainability and privacy). The proof-of-concept system built was designed to: [28]

- Deliver broadcast messages from network providers to OBEs at specified geographic locations
- Deliver broadcast messages from local systems such as traffic signals or toll stations to OBEs at specified geographic locations
- Deliver broadcast messages between OBEs
- Collect data from OBEs and distribute topical information extracted from the data to network subscribers

- Provide OBEs access to remote private service providers, and this access can be carried over from one RSE to the next without disrupting the service
- Provide security functions to protect against attacks and to protect the privacy of the individual users.

The VII initiative was successful in demonstrating a prototype system capable of delivering most of the stated objectives. Most importantly (and relevant to this thesis), this project had resulted in the development and eventually the rectification of the Dedicated Short Range Communication (DSRC) group of protocols. The project also revealed shortcomings in the technologies used, including concerns about the reliability of antennae, communication protocols and positioning accuracy.

Following on from the VII project, in order to address the shortcomings identified and to implement a more up-to-date vision of the next-generation road network (the "Connected Transportation Environment"), the USDOT commissioned further research for the five years from 2010 to 2014 with the aim of furthering the development on V2V and V2I technologies [3]. Relevant to this thesis is the work on "connected vehicle", where the outcomes from the VII study are being field-trialled.

Similar to USA, European countries were also developing a parallel set of technologies for the next-generation ITS. A number of projects were commissioned by the EU in the early 2000s looking into adding data communication support into vehicles. The Car 2 Car Communication Consortium (C2C-CC), eSafetyForum, ISO/CEN project on "Communications Access for Land Mobiles" (ISO-CALM), *etc.* were formed, investigating various aspects of vehicular communications.

Currently, the European Telecommunications Standards Institute (ETSI), together with the European Committee for Standardization (CEN), is developing an integrated and globally compatible structure of the entire ITS Communication (ITSC) architecture, in addition to the standards for the individual components.

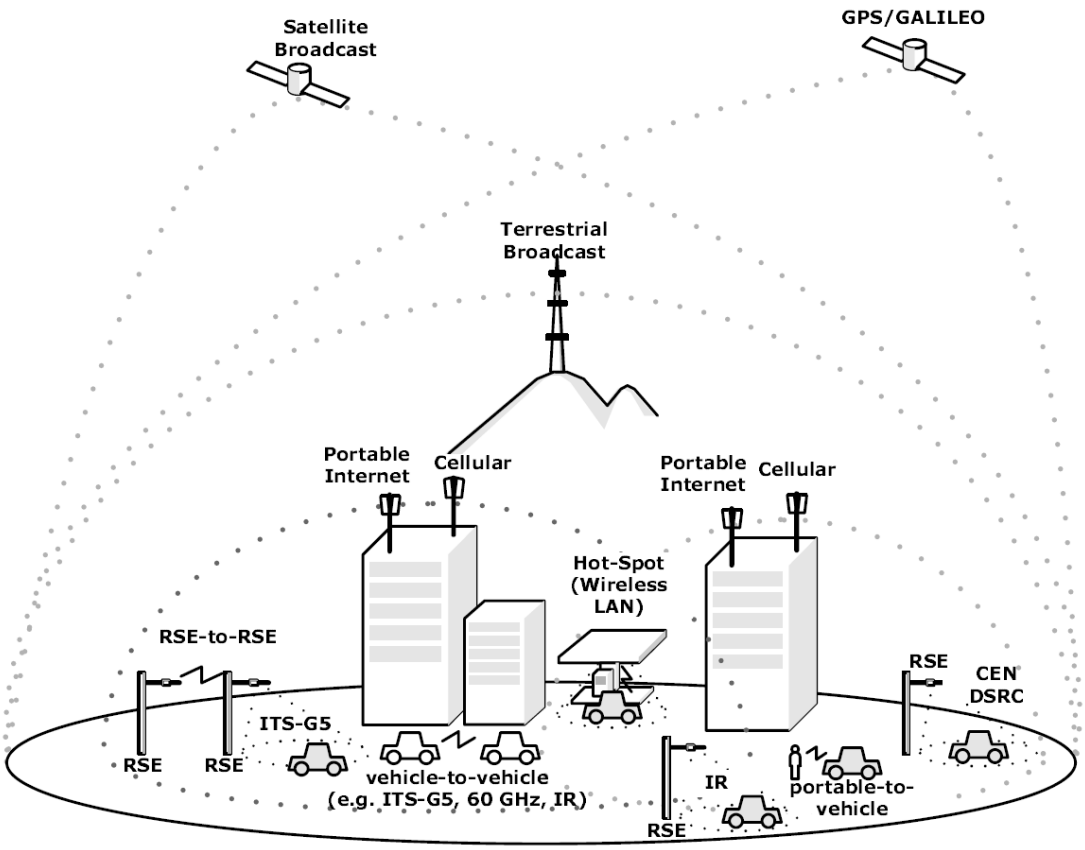


Figure 2.3: ETSI envisaged mode of ITS communications [32, Figure 1]

The overall architecture of ITSC is defined in ETSI EN 302 665 [32]. Figure 2.3 depicts the envisioned communications pattern for the ITSC infrastructure. In that vision, the components of the ITS may communicate with each other using different media depending on the usage scenario. The standards were developed to be able to support a variety of underlying physical technologies, ranging from physical wire and infrared, to satellites and DSRC (*a.k.a.* "ITS-G5", "CALM M5"). ETSI standards define the overarching architecture, integrating the previous work from the various projects, including the large-scale multi-modal communications in the ISO-CALM project, system designs from C2C-CC, as well as the physical and MAC layer specifications from IEEE.

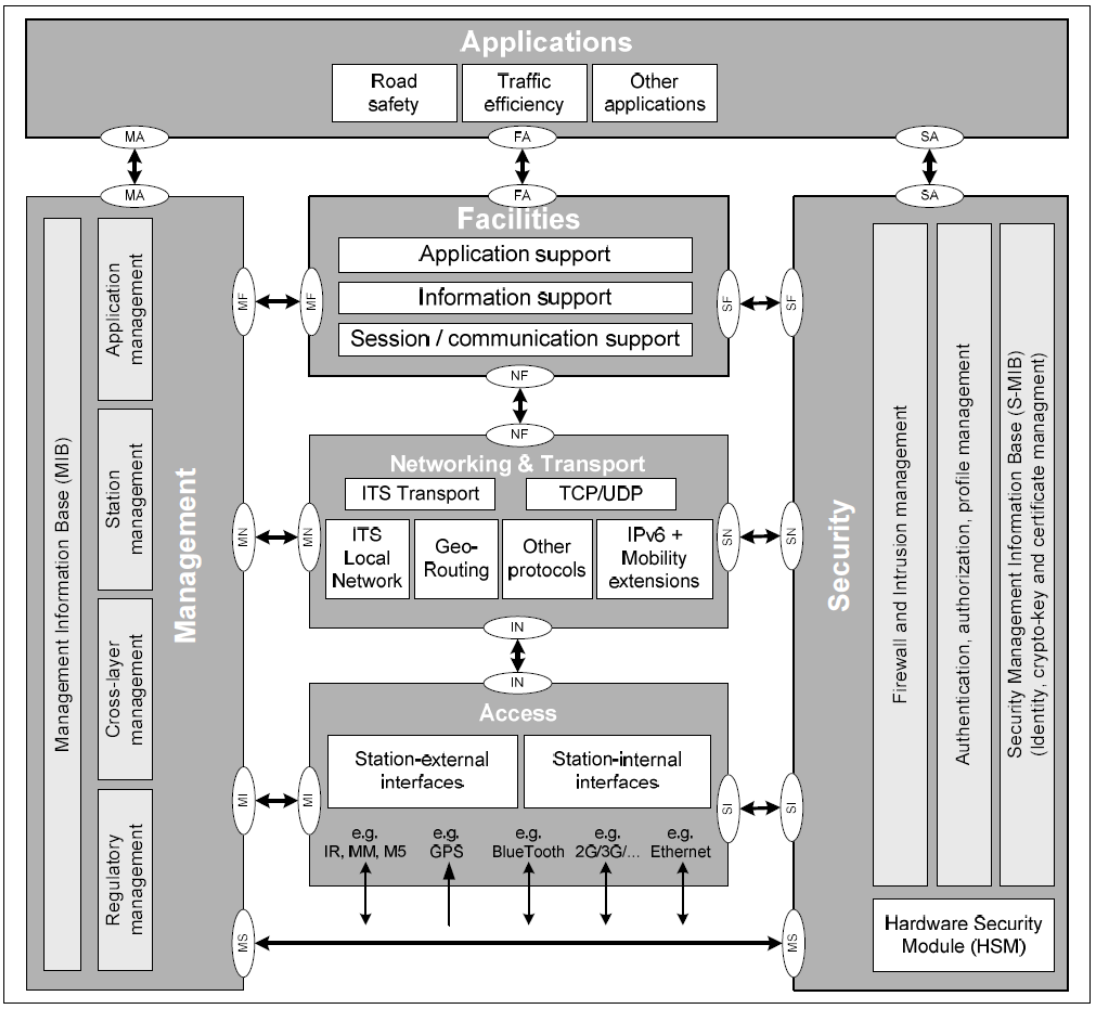


Figure 2.4: ETSI ITSC station architecture [32, Figure 3]

2.4.1 Communication protocol stack

The overall communication protocol stack used in ITSC is being harmonised across the world, with the harmonisation efforts for the majority of the lower layers already completed. The discussion below is based on the harmonised ETSI model, with differences between American and European models highlighted.

In general, the protocol stack used for ITSC is relatively simple in order to minimise processing latency. The high-level view of the protocol stack (from the ETSI model) features three protocol layers and two cross-layer functional planes (Figure 2.4). Note — the "application" block in the figure represents the actual applications using the stack, not the OSI Application Layer (layer 7).

The higher functional layers (OSI layers 5, 6, and 7) are defined in the "Fa-

cilities" layer. This layer defines the message formats and processing for various application communication requirements. Implementations of these functions are driven by the Society of Automotive Engineers (SAE) in USA and ISO-CALM in Europe. Most of the work in this layer are to be harmonised. The SAE J2735 [33] "Dedicated Short Range Communications (DSRC) Message Set Dictionary" is a standard developed by SAE specifically for ITSC, version 2 of which will contain the harmonised set of messages.

The "Network & Transport" layer in ETSI architecture deals with the (potentially) reliable routing, *i.e.* OSI layers 3 and 4. The architecture is designed to support both TCP/UDP over IPv6 (with optimisations), as well as ITSC-specific protocols. This layer implements the geographical routing and broadcasting functions, which are essential for many VANET applications, and is relevant to the work in this thesis. It is noted that the American standards initially did not specify any routing strategies, leaving routing and reliability to "higher layer application" definitions (Figure 2.5). The European solution integrates the network layer from C2C-CC, enabling wireless multi-hop forwarding based on geographical addressing and routing. It implements location table, beaconing, location service, and geographical addressing and forwarding algorithms, as well as congestion control (Figure 2.6).

The actual communication technology (OSI layers 1 and 2) is specified in the access layer of the ETSI architecture. The European model specifically supports a wide range of access technologies, including infrared, Bluetooth, GSM / UMTS / LTE mobile data, LAN/WLAN, as well as ITS-G5 (*i.e.* DSRC in American terminology). DSRC is defined in IEEE 802.11p and IEEE 1609.4 (Figure 2.5). IEEE 802.11p is an extension to the IEEE 802.11 PHY and MAC layers, while IEEE 1609.4 specifies channel use. This thesis focuses on the DSRC (ITS-G5) technology — the other access methods will not be further discussed.

Multi-DSRC channel access is specified in IEEE 1609.4-2010 [36], detailing the higher layer functions when the vehicular network operates on multiple chan-

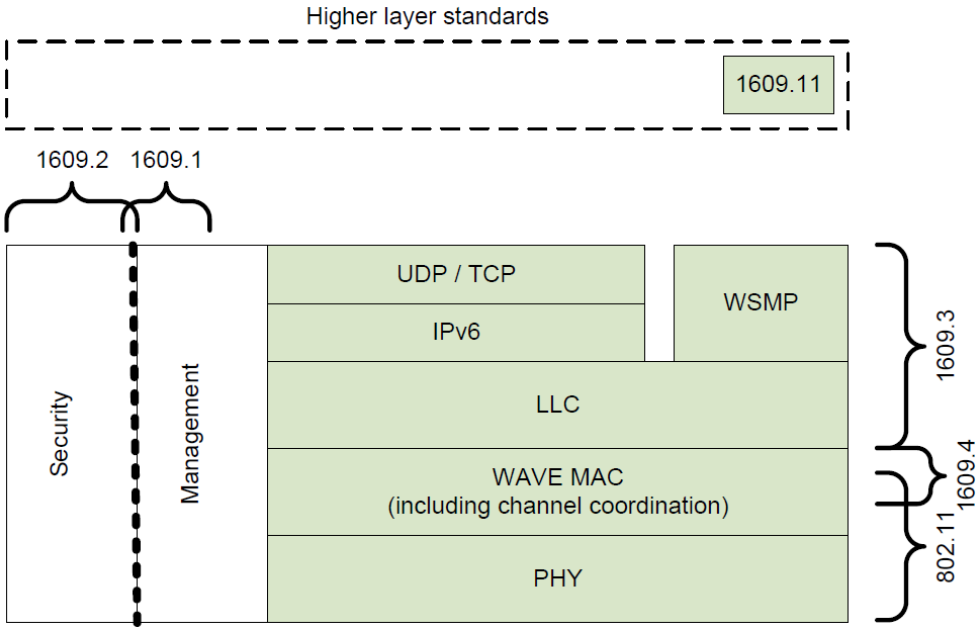


Figure 2.5: IEEE DSRC/WAVE protocol architecture [34]

nels. This standard allocates some time period (the CCH Interval) during which all stations must switch to and monitor a control channel (CCH), while at other times (the SCH Interval) the station may operate on service channels (SCH) (Figure 2.7). This has the net effect of reducing the total capacity of the control channel as the CCH is only guaranteed to be monitored for a subset of the time, and also increases channel contention at the start of the CCH interval. Studies have presented analysis and simulation results demonstrating the limitation introduced by IEEE 1609.4 [37–40], thus the efficient use of the CCH interval is crucial to the success of DSRC-based VANETs. Discussions and analysis in this thesis concentrate on the MAC layer and network layers, and do not consider the effects due to multi-channel operation.

In addition to the stack of network protocols, two cross-layer planes were also defined. The ESTI architecture contains a Management plane that deals with the overall management of the device, as well as ensuring regulatory conformance. The other is the Security plane, involved in coordinating privacy and security safeguards in the station. These two planes are still being harmonised. In USA, the Management plane is specified in IEEE 1609.1, while the Security plane is

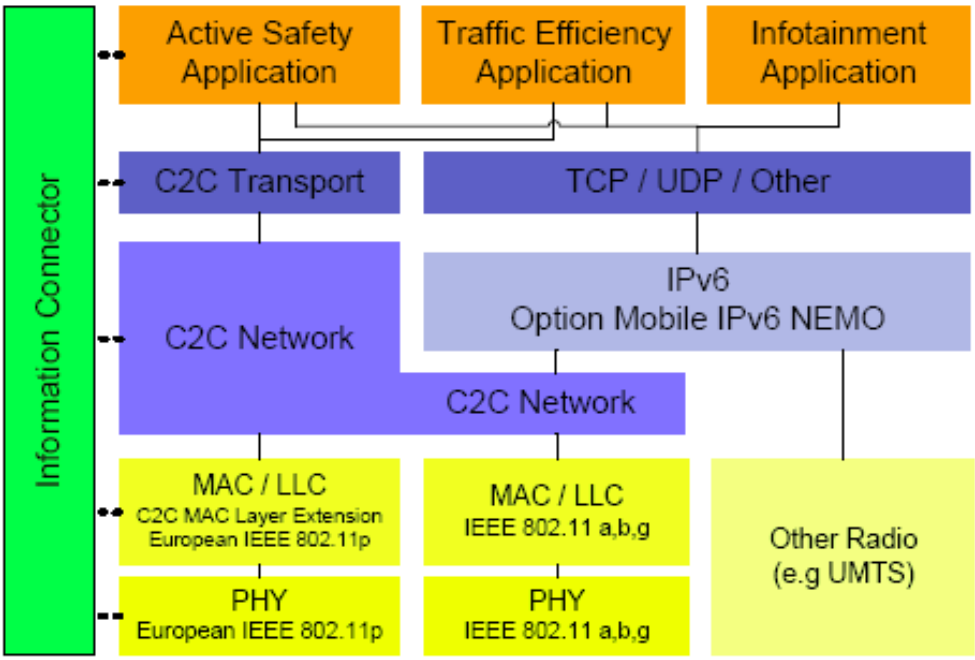


Figure 2.6: C2C-CC networking design [35]

specified in IEEE 1609.2 (Figure 2.5).

ETSI also developed an outline of how the sub-systems of the ITSC network interact with each other. Figure 2.8, taken from the ETSI document [32], shows how various components in the network may implement different subsets of the protocol stack, and how the components may be connected to each other. For example, a vehicle ITS sub-system may implement a full stack in order to utilise the ITS services, a gateway that connects to the vehicle's internal control network,

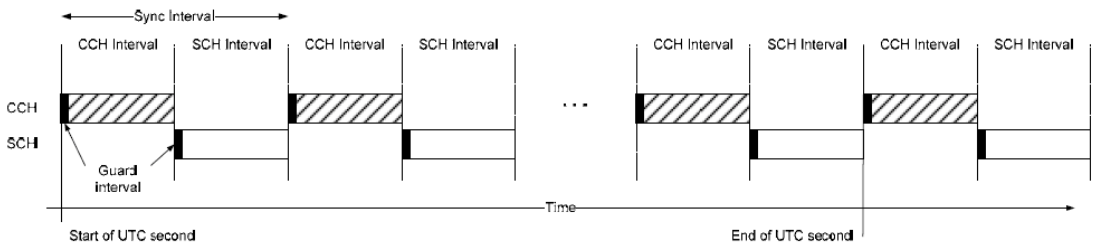


Figure 2.7: IEEE 1609.4 sync interval, guard interval, CCH interval and SCH interval [36, Figure 4]

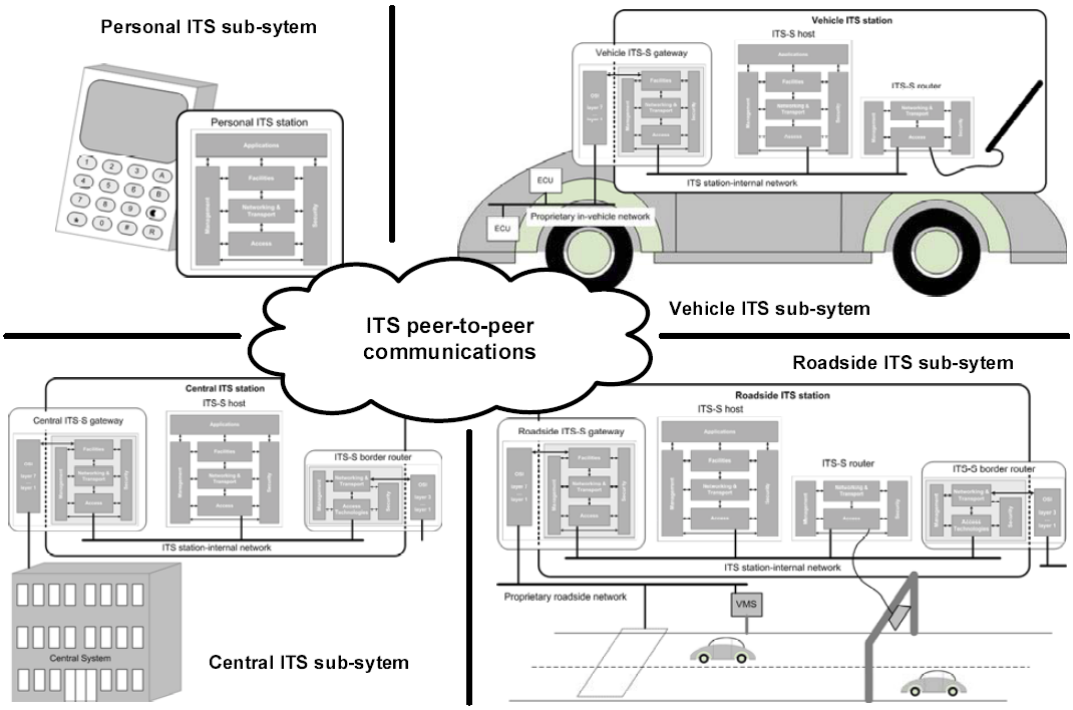


Figure 2.8: ITSC subsystem interaction [32, Figure 4]

and separately, a router sub-system partially implement the stack and connects to the outside network. The separation of the router sub-system from the control network enhances isolation of core functionalities, minimising the impact of potential security breaches.

2.4.2 Dedicated Short Range Communications (DSRC)

DSRC (also known as ITS-G5 and CALM-M5) is the central feature of V2V communication. DSRC operates at 5.8-5.9 GHz, a slice of the radio spectrum dedicated to vehicular communications, using OFDM and with a bandwidth of 10 or 20 MHz per channel (depending on the regulation and applications).

The DSRC network protocols are specified in IEEE 802.11 and its amendment IEEE 802.11p. The changes introduced by IEEE 802.11p enables "association-less" mode of operation, removing the requirement for stations to transmit/discover group membership status before operating on the channel. This change overcomes the problem with the very high station mobility, with many station

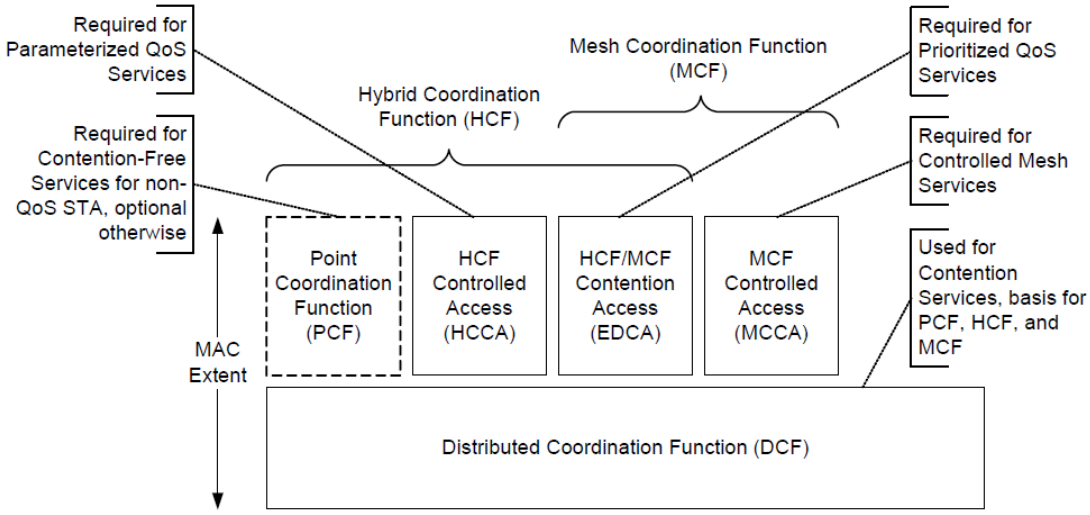


Figure 2.9: IEEE 802.11 MAC architecture [44, Figure 9-1]

joining and leaving within very short time frames. In summary, stations communicate following the standard CSMA/CA procedure as specified in IEEE 802.11, with binary exponential backoff retransmissions and optional RTS/CTS for unicast. IEEE 802.11p uses the normal IEEE 802.11 frame formats. The typical use-case for DSRC safety applications is broadcast-mode transmissions, which are not often used in WLAN for data packets. A good understanding of the IEEE 802.11 backoff process is essential in order to effectively use its properties to estimate channel load.

2.5 MAC-layer broadcasts in VANETs

The medium access (MAC) protocol used in DSRC-based VANETs is described in IEEE 802.11 with amendments. IEEE 802.11p [41] incorporates the physical layer (PHY) timing parameters from IEEE 802.11a [42], the Quality-of-Service (QoS) mechanisms in IEEE 802.11e [43], amends PHY parameters (frequencies, bandwidths, *etc.*), and enables association-less communications. These changes are essential to DSRC, but do not change the behaviour of MAC broadcasts and will not be further discussed. All changes specified in IEEE 802.11p are now incorporated into the revised standard IEEE 802.11-2012 [44].

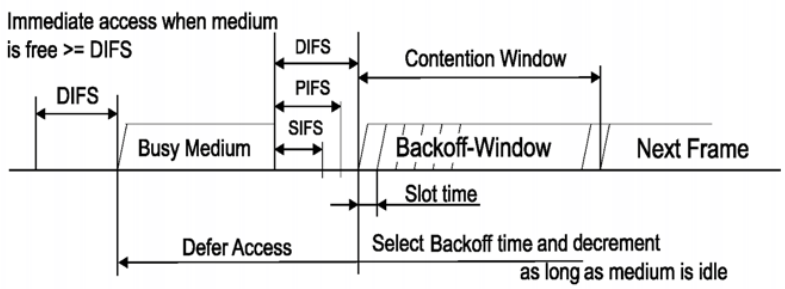


Figure 2.10: IEEE 802.11 basic access method [44, Figure 9-11]

The IEEE 802.11 standard provides two means to access the channel relevant to non-mesh operations — contention-based channel access during the "Contention Period" (CP) where stations compete for the channel in a distributed manner; and the contention-free "HCF controlled channel access" during an optional "Contention-free Period" when the "Hybrid Coordinator" (*e.g.* a base station) polls all stations and time-schedules their use of the channel. Furthermore, broadcasts during the CP may or may not be facilitated by a base station (Figure 2.9). Only the contention-based channel access not facilitated by a coordinator (typical for VANETs with rapidly changing network topologies), and in the absence of station clustering schemes (*i.e.* no obvious coordinators), is considered in this thesis.

Channel access during the Contention Period is governed by the Enhanced Distributed Channel Access Function (EDCAF). Non-QoS enabled stations operating under EDCAF behaves in the same way as those under the Distributed Coordination Function (DCF). (This thesis may refer to EDCAF and DCF interchangeably.) In the non-QoS case, each station maintains a single backoff counter, initialised to a value uniformly chosen within the initial contention window $CW = [0, aCW_{min}]$. When the medium is sensed to be idle, stations having something to transmit will not send the packet immediately, but will continue to wait for a short time called an "Inter-Frame Space" (IFS). If the medium is still idle after the IFS and the backoff counter is zero, the frame at the head of its transmit queue is then transmitted. If the backoff counter is non zero after the

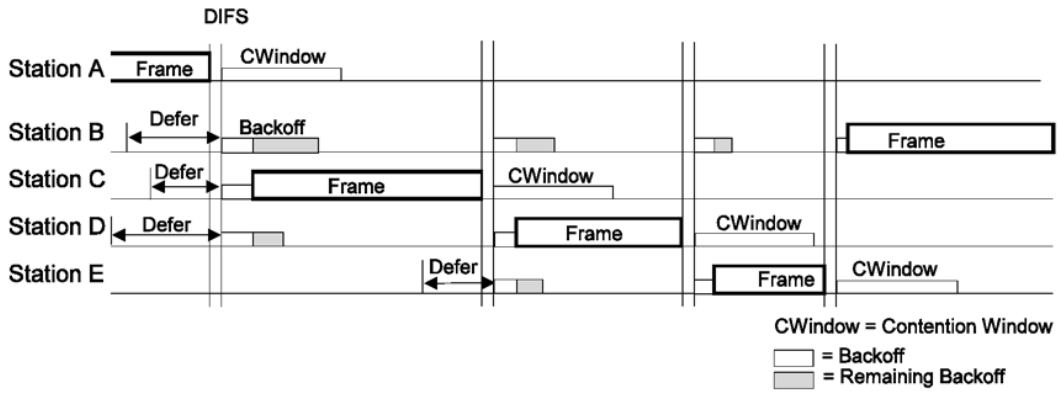


Figure 2.11: IEEE 802.11 backoff procedure [44, Figure 9-12] — When a station has something to transmit, it first senses the channel. If the channel is busy, it defers access until a period equals to DIFS after the channel becomes idle (Stations B, C, D). The station then begins decrementing its backoff counter. The first station with counter equals to zero is permitted to transmit (Station C). Upon the channel becoming busy again, all other stations suspend decrementing their backoff counters until DIFS after the channel becomes idle (Stations B, D).

IFS, as long as the medium remains idle, the backoff counter is decremented after a slot time (ST) and the station again checks the value of the backoff counter. The station will continue decrementing the counter every ST as long as the medium remains idle. The medium becomes busy when any station (including the station itself) starts transmitting, in which case the backoff counter pauses until the medium becomes idle again and the process is repeated. After a frame has been transmitted by the station, the backoff counter is reset to a value within the initial CW. The chosen value should be uniformly distributed, but some vendors are known to violate this specification [45]. For unicast messages, an Acknowledgement (ACK) frame will be sent by the receiver to indicate correct receipt of the frame. This channel access scheme is called "Basic Access". Figures 2.10 and 2.11 illustrate the Basic Access scheme and the backoff procedure. In addition, unicast messages may also be preceded by a "request to send" (RTS) and "clear to send" (CTS) handshake, designed to activate "virtual carrier sense" to mitigate hidden terminal problems. For broadcast messages, because there are multiple unspecified and possibly unknown receivers, both the use of ACK and RTS/CTS do not make sense, and therefore is disallowed in the standard [44, Section 9.3.6].

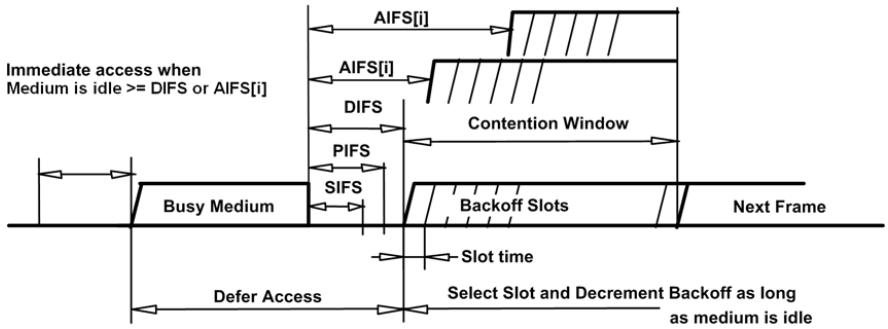


Figure 2.12: IEEE 802.11 IFS relationships [44, Figure 9-3]

The RTS/CTS procedure will not be further considered in this thesis.

There are also different types of frames defined in the IEEE 802.11 standard, some with higher priority than others (*e.g.* control frames takes precedence over data frames). In order to prioritise these frames, the standard specifies different IFS — Reduced IFS (RIFS), Short IFS (SIFS), PCF IFS (PIFS), DCF IFS (DIFS), Arbitration IFS (AIFS) and Extended IFS (EIFS) (see Figure 2.12). These IFS define the length of time between the channel is sensed to be idle following a transmission, and the station is permitted to transmit a frame or to start decrementing their backoff counters. IFS are typically of different durations, and therefore, with the exception of RIFS and EIFS, specifies the priority of frames that are sent. (Frames that uses a shorter IFS has a higher priority to ones that uses longer IFS.) The RIFS is designed to replace SIFS to reduce waiting time when SIFS-spaced frames are not expected (*e.g.* no ACK frames expected for broadcasts). The EIFS is another special IFS that is not used for prioritising frames, but is used to prevent a frame being transmitted over an ACK frame from a hidden terminal. EIFS is triggered when a station cannot successfully decode a frame (*i.e.* a station sensed the channel to be busy, but the frame header was decoded with errors.) For heavily congested channels, it is likely that EIFS may frequently be triggered due to frame collisions, which EIFS not only cannot help mitigate, but also act to further reduce the channel capacity. When making observations on the channel contention, it is important to take into account the effects EIFS may have on the observed value.

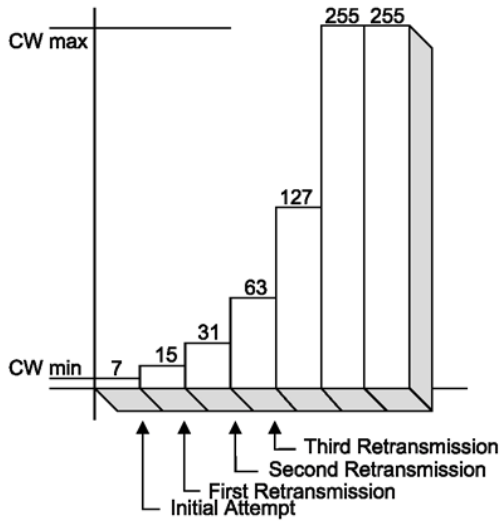


Figure 2.13: Example of exponential increase of CW [44, Figure 9-10]

In order to balance channel utilisation and collision probability (hence throughput), the EDCAF (and DCF) adjusts the contention window size dynamically for unicasts. A small CW improves utilisation when contention is low as stations do not have to wait as long before transmitting, whereas a large CW reduces the chance of multiple stations choosing the same backoff for transmission (*i.e.* collisions). A "binary exponential backoff" mechanism (Figure 2.13) is used in unicasts to increase the contention window when the channel is congested. Both EDCAF and DCF determines channel contention by assuming frames are lost only due to packet collisions, which becomes more frequent as channel contention increases. Therefore, when frames are lost (indicated by the non-receipt of an ACK frame), the contention window size is doubled until reaching aCW_{max}, a parameter specified in the standards. In broadcasts, no feedback mechanism is available to indicate the success or failure of transmissions, thus the exponential backoff is never invoked and the CW always remain at $[0, aCW_{min}]$. As a result, the IEEE 802.11 EDCAF is not able to adapt to channel contention in cases where a large proportion of transmissions are broadcasts, as can be expected on the control channel of VANETs.

The IEEE standards for vehicular networks also incorporates support for Quality-of-Service (QoS) mechanisms. In IEEE 802.11 four classes of priorities

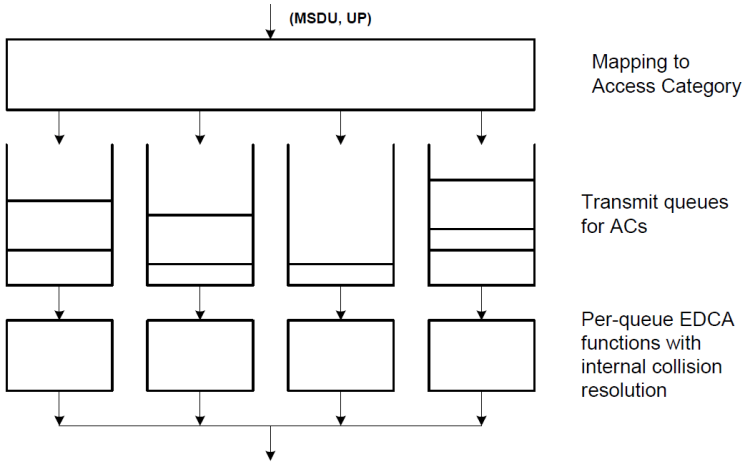


Figure 2.14: IEEE 802.11 reference implementation model [44, Figure 9-19]

(called "Access Categories" or ACs) are defined (Table 2.3) — voice (AC_VO), video (AC_VI), best-effort (AC_BE) and background (AC_BK). QoS is provided by putting frames of different categories in their own separate queue. These queues operates as though they are separate stations, individually participating in the CSMA/CA process and maintaining their own backoff counter. The standard defines different values of AIFS, min and max CW size for the different queues, thereby achieving a probabilistic service differentiation. There is a separate mechanism defined for the case where a "virtual collision" occurs between the different queues within the same station — the higher priority frames is transmitted, and the lower priority frames will follow exponential backoff. (Figure 2.14) The work in this thesis does not consider network QoS, therefore their effects

Table 2.3: IEEE 802.11 UP-to-AC mappings [44, Table 9-1]

Priority	UP (Same as 802.1D user priority)	802.1D designation	AC	Designation (informative)
Lowest	1	BK	AC_BK	Background
	2	—	AC_BK	Background
	0	BE	AC_BE	Best Effort
	3	EE	AC_BE	Best Effort
	4	CL	AC_VI	Video
	5	VI	AC_VI	Video
	6	VO	AC_VO	Voice
Highest	7	NC	AC_VO	Voice

on channel saturation behaviour will not be further discussed. There are other analytical works such as [46–49], which may be adapted in a similar fashion to the work in Chapter 5 to allow the techniques presented in this thesis be used.

Overview

This chapter is a literature survey on existing work related to the investigations in this thesis.

Chapter 3

Related works

3.1 Improving packet reception

VANETs, being wireless communication systems, are exposed to many effects that can limit their reliability, and therefore limiting the reliability of any safety systems based on them. Since VANET stations are highly mobile, they may at times experience significant multipath fading (*e.g.* in tunnels) and shadowing (*e.g.* behind heavy vehicles or buildings), causing packet loss. The U.S. VII Consortium, in their final report, found that while the DSRC transmission ranges between tall vehicles are adequate, the effective range amongst low vehicles are inadequate, with shortcomings evident in all proof-of-concept test scenarios [28].

Stanica *et al.* presented investigations on the reasons for broadcast packet losses in VANETs [50]. They have identified radio propagation problems, the time-sensitive nature of VANET broadcasts, packet collisions and hidden terminals as the main causes of packet loss. They noted that the severity of link layer packet loss can be controlled by adjusting beaconing frequency, PHY/MAC layer data rate, transmission power, contention window size and the carrier sense threshold.

Even though the IEEE 802.11 standard specifies automatic repeat request (ARQ) techniques for error recovery, they cannot be used because VANET messages are typically sent in broadcast mode. While occasional packets losses may

be acceptable for "comfort" applications, they are detrimental to safety systems such as Cooperative Collision Avoidance (CCA) [22] and Post-Collision Warning (*a.k.a.* Cooperative Collision Warning, CCW) [25]. Packet loss reduces the accuracy of these systems, either resulting in dangerous scenarios not being detected and reacted upon, or in users no longer trusting these systems. Bastani *et al.* found that the hidden terminal problem is a significant contributor to packet loss in VANETs, especially impacting high speed road segments [51]. Ma *et al.*, using mathematical modelling, had identified that the standard IEEE 802.11 mechanisms are insufficient for providing the required packet reception ratio, even when packet losses are caused only by MAC-layer mechanisms [52].

Previous experiments and simulations [25, 53] have identified that a large proportion of transmitted packets are lost due to fading [53], even when the receiver is as close to the transmitter as half the maximum transmission radius. In contrast, Bai and Krishnan conducted physical experiments investigating the effect of distance on packet delivery ratio, and found that in the absence of obstructions, packet reception are acceptable, with minimal consecutive packet losses in highway settings [54]. Their findings suggest that, in the absence of obstructions, the fast fading effect dominates in highway situations, but the magnitude of the fast fading appears to be lower than predicted under the Nakagami model used in previous simulation studies. However, obstructions between the transmitter and the receiver were shown to further attenuate the signal, thereby decreasing the reliability of the system.

It is well known that heavy vehicles cause shadowing in radio channels. Radio channel measurements and analyses conducted for mobile telephony in the 1990s [55] showed that large vehicles can cast deep shadows between 6 dB and 30 dB at 900 MHz in an experiment that measured signal strength across the width of a tunnel. Rustako *et al.* measured an attenuation of up to 20 dB at 11 GHz when using monopole antennae, and based on their model, an attenua-

tion of up to 20 dB at 4 GHz can be expected [56]. In a real world study, Klingler measured an additional attenuation of 50 dB over 3 km when heavy vehicle flow is at 150 trucks per hour, compared to 2 trucks per hour over the same distance of road [57].

More recently, radio channel measurements were also conducted specifically for DSRC applications. Paier *et al.* reported field measurements that revealed non-trivial delay-spread and Doppler shifts in addition to shadowing loss of approximately 20 dB across trucks at 5.6 GHz [58]. Meireles *et al.* also observed an attenuation of up to 20 dB across small vans, and up to 27 dB across semi-trailers in static environments (parked vehicles) [59, 60]. Those experiments also revealed significant deterioration in packet delivery ratio in on-road scenarios. Abbas *et al.* presented a channel model in attempt to characterise the effects of vehicle obstructions [61]. In their measurement study, Abbas *et al.* differentiated between line-of-sight (LOS), line-of-sight obstructed by vehicles (OLOS) and line-of-sight obstructed by buildings, and found that the LOS and OLOS scenarios can be modelled using log-normal distribution, with means separated by 10 dB, implying a mean attenuation 10 dB can be attributed to vehicles' shadows.

In order to combat packet loss resulting from this shadowing effect, hence increasing the robustness of VANET transmissions, this thesis presents a load-adaptive and interference-aware geocast system. Even though many work have been proposed for improving packet reception, they often introduce excessive interference into the system, thereby causing unintentional packet collisions.

Existing methods to improve packet reception ratio fall into two categories: localised strategies that operate only on the sender, and forwarding-based strategies that operates across all stations on the network.

3.1.1 Localised strategies

Localised strategies for improving packet reception include QoS-based techniques [24, 53] and repetition techniques [22]. They tend to generate lower inter-

ference than forwarding-based techniques, but are not very effective in the presence of shadowing.

QoS-based techniques work by giving priority to emergency traffic, thus reducing the channel contention for this class of traffic. Torrent-Moreno *et al.* investigated the effect of using IEEE 802.11e QoS technique to prioritise warning messages [53]. By placing warning messages in a higher access category than routine traffic, Torrent-Moreno *et al.* had demonstrated a higher packet reception ratio for the warning messages, at the cost of lower reception ratio for routine traffic. Unlike Torrent-Moreno *et al.*'s work, Biswas *et al.* investigated the effect of a local QoS scheme where each station always schedules CCW (denoted as "CCA" in the paper) packets to transmit in preference to any routine packets from that station [24]. They have shown that this type of QoS scheme can also boost reception of warning messages. Wischhof and Rohling [62] investigated a QoS scheme whereby each station prioritises its traffic based on some "utility" metric, and implements the station's transmit queue as a finite priority drop-tail queue based on that metric such that lower priority packets are discarded as necessary. All the techniques above improved reception ratio and/or delay for a subset of packets, but either reduce reception ratio, or increase delay for the remainder.

Unlike QoS-based techniques, repetition techniques improve packet reception by adding temporal redundancy, thus helping to overcome fast fading and packet collisions. Xu *et al.* tested a number of repetition techniques with simulations in scenarios with no fading [22]. They have tested cases where stations are or are not synchronised, CSMA is or is not followed and the number of repetitions is either fixed or p-persistent (repeat with certain probability p). Xu *et al.* found that both having a fixed number of retransmissions and the use of CSMA produce the best results, but noted that these repetition techniques are ineffective for mitigating packet losses due to shadowing [22].

Artimy *et al.* presented an interesting alternative technique for controlling

channel contention — adjusting transmission power. The technique control channel contention by moderating the amount of stations affected by another station's transmission, with lower transmission power affecting only the stations that are closer to the transmitter. They presented a method to estimate the local station density based on traffic flow theory, and adjusted transmission power based on the density estimate [63, 64]. Torrent-Moreno *et al.* used transmit power adaptation technique for another effect — packet capture [65]. Their central idea is that packet collisions are not necessarily bad because the colliding packet with higher receive power can often be recovered. By carefully considering the transmit power, two stations can concurrently transmit with the packets being received correctly by stations closer to the transmitters (hence probably more relevant). Sebastian *et al.* [66] also looked into adjusting transmission power, but unlike Torrent-Moreno *et al.*'s work, Sebastian *et al.* only used this technique as part of a routing algorithm. Many other works have since been published describing methods to improve transmission power control, with some requiring explicit coordination amongst vehicles, and others by inference from channel observations.

Recent work by Stanica *et al.* also tries to exploit packet capture effect [67]. In Safety Range Carrier Sense Multiple Access (SR-CSMA), instead of exploiting packet capture by adjusting transmission power similar to Torrent-Moreno *et al.*'s scheme [65], they adjust the carrier sense threshold. By increasing the carrier sense threshold, their technique ignores concurrent transmissions from farther stations, thereby increasing the collision probability with those stations. The authors reasoned that collisions with packets from outside the safety range doesn't matter as long as the packets from within the safety range can be reliably received.

In the physical layer, there had also been work aimed at reducing contention. Chigan and Li presented a technique that uses directional antennae (or antenna arrays) in order to minimise the interference introduced [68]. Such technique is not currently feasible due to cost. Furthermore it is also well known that the use of directional antennae can adversely affect the CSMA/CA protocol used

in DSRC — directional antenna reduces the interference to non-target stations, which also means that those stations cannot detect a concurrent transmission. This introduces extra hidden-terminal problem into the already shadow-prone channel.

Another method to reduce channel contention is to improve the determination of the target area. By being more precise in determining and specifying the target area, in combination with forwarding-based techniques, the interference to stations outside the more refined target can be reduced, hence reducing overall channel contention and collision, and improving reception. Xian and Huang [69] proposed a method that relies on a map database to help determine the target area of warning messages. The technique of better specifying targets operates orthogonally to other techniques for improving reception, and will not be further discussed in this thesis.

While these localised techniques have each been shown to improve packet reception, as explained by Xu *et al.* [22], they are ineffective in overcoming packet loss caused by shadowing.

3.1.2 Forwarding-based strategies

Unlike localised techniques, forwarding-based strategies exploit spatial diversity to route around obstacles. These can overcome problems caused by shadowing if suitable relays exist. Multi-hop message forwarding typically belongs in OSI Layer 3 (Networking), and many routing and cooperative retransmission schemes have been proposed for unicast in wireless ad hoc or mesh networks with the aim of minimising the number of hops and/or latency. This thesis concentrates the discussions on broadcast and geocast algorithms, some of which were adapted from unicast work. Maihöfer [70] published a comprehensive survey of geocast algorithms in 2004, which has since been extended to address VANET-specific issues by Li and Wang in their well cited paper [71], and subsequently by Chen *et al.* in [72].

In terms of geographically-bound broadcasting (geocasting), there are three main classes of techniques — predetermined routes (*e.g.* by routing table or just-in-time route discovery, *etc.*), clustering and flooding. Techniques using predetermined routes are typically more optimal (in whatever metric the algorithm was optimised for), but requires some knowledge of the network topography. Flooding-based techniques often require minimal *a priori* knowledge, but may be sub-optimal and potentially causes many irrelevant/redundant packets to be transmitted due to the lack of coordination between stations. There are also other techniques (*e.g.* data aggregation, packet concatenation and piggybacking) designed to operate at higher communication layers in order to lessen the load offered to the network.

Predetermined routes

The predetermined route methods involve building and updating routing tables and/or just-in-time route discovery. In networks where the topology is known, there exist algorithms that can achieve high reception ratios (deterministically or probabilistically) at relatively low communication costs. Peng and Luo [73] described an algorithm where each station knows the exact network topology, and waits briefly before forwarding to its neighbours. During the delay, the relays individually keep track of retransmissions overheard, and determine whether it still needs to retransmit. If all its neighbours had been covered already, the relay will not forward the packet. Unfortunately, the requirement of perfect 2-hop knowledge makes this algorithm unsuitable for ad hoc networks.

Ros *et al.* transformed the problem of broadcast into a graph-theoretic problem of determining a connected dominating set (CDS). In their algorithm [74], VANET broadcast/geocast is achieved by a series of unicasts to members of a CDS. Ros *et al.*'s work provides a method to maintain and update neighbourhood information by recording broadcast message received and by non-reception of ACK packets. Their algorithm abstracts out the technique of determining a

CDS, which the author admitted is NP-complete, simply stating that "computing a CDS in a VANET environment comes for free" due to the pervasive position information of all stations on the network. The use of multiple unicasts to achieve broadcast and the complexity of determining a CDS especially when the topology is constantly changing make such scheme difficult to implement for time-sensitive messages.

Sebastian *et al.* investigated the issue of localised interference (*i.e.* local channel contention) during geocasts [66]. In their graph-theoretic work, Sebastian *et al.* reduced the problem of geocast to a vertex-weighted delay-constrained minimum Steiner tree (which is also NP-complete), with vertex cost being the area a relay station covers at a given transmission range. Each station may change its transmission range to cover more or less stations. This algorithm produces a theoretically good set of relays that satisfies given delay requirements, achieving good coverage while minimising interference. Unfortunately, similar to Peng and Luo's [73] and Ros *et al.*'s [74] works, the need to first transform the problem into a graph is problematic because the stations in VANETs are highly mobile, thus requiring very frequent updates or else the cached graph would be stale, making this technique not very suitable for VANETs.

In ad hoc networks such as MANETs and VANETs, station connectivity changes frequently, making the maintenance of complete network knowledge not scalable. Direct Source Routing (DSR) [75] and Ad hoc On-Demand Vector Routing (AODV) [76] are two protocols designed to discover and maintain partial knowledge of the network topology on demand, thereby reducing the cost involved in maintaining relevant network knowledge.

In Direct Source Routing (DSR) [75], the source station floods a Route Request to the destination only when it has something to send and the source does not have a valid cached route entry. The source and each intermediate station append its identity onto the routing path field. When the destination receives the request it sends a Route Reply back to the source, either by specifying the

reversed routing path in the IPv4 source route field, or by flooding a Route Request packet back to source with a piggybacked Route Reply. Once the route had been discovered, messages are sent between the source and destination by source routing. There are also mechanisms in DSR to recover broken routes. In this scheme, only the source and the destination need to maintain some partial knowledge of the network. Unfortunately, while DSR can be used for unicast messages, broadcast/geocast packets cannot be source routed and therefore cannot utilise DSR.

Ad hoc On-Demand Vector Routing (AODV) [76] is another routing protocol designed for MANETs. Similar to DSR, AODV message source floods a Route Request (RREQ) packet across the network only when it has something to send. As destinations are reached, Route Reply (RREP) packets are sent back via the original route. Unlike DSR, AODV does not use source routing. Instead, intermediate stations between source and destination aggregates all the RREP received to compile and maintain a partial routing table for the communicating stations, and can easily support broadcasts and multicasts. AODV had also been extended to support geocast, for example, in Context-Based AODV [77, 78] where RREQ packets also incorporate details of geographical zones. AODV is able to maximise packet delivery while minimising redundancy but require routing tables to be continuously updated, and may also cause packet losses if intermediate stations leave the network before the routing table is updated. Wang *et al.* conducted an experiment using 6 sedan vehicles, and found that AODV was unable to find and maintain long routes, and suffers excessive packet losses [79]. Furthermore, DSR- and AODV-like algorithms require a route establishment phase (RREQ-RREP exchanges), which is unsuitable for VANET safety messages as the handshake may require delays of more than 200 ms [24].

In a scheme similar to DSR, Liu and Seet *et al.* demonstrated the potential improvements that can be attained by incorporating knowledge of the environment in their A-STAR protocol [80, 81]. In A-STAR, stations rely on matching

statistically related maps to determine the optimal "junctions" (relay positions), and then follows geographical source routing. The authors demonstrated a 40% improvement in packet delivery when (vehicular) traffic-aware A-STAR was used for unicasts.

Flooding

Because route establishment and recovery is costly in terms of time, it can be argued that flooding, which does not need to maintain states, may be a more suitable method for the dissemination of VANET safety messages. Flooding-based schemes typically assume no knowledge of network topology, and can use only relatively local knowledge to make forwarding decisions. Since stations do not coordinate amongst each other, transmissions might be triggered at suboptimal locations, causing stations to receive more duplicate and/or irrelevant packets than in other methods. This therefore generates a larger amount of interference both within and outside the coverage area. Ni *et al.* showed that packet retransmissions in MANETs are highly redundant, with relays capable of providing only up to 61% additional coverage under unit-disc coverage assumptions [82,83]. This interference reduces the capacity of the network both by taking up air time as well as causing collisions with other transmissions. When broadcast messages are infrequent, these schemes can function relatively well, but they are unable to handle the high broadcast loads as expected from CCA systems.

Ni and Tseng *et al.* [82, 83] described and compared five classes of techniques to reduce the cost of broadcasting in MANETs. First, probabilistic forwarding, which is also known as "gossipping", modifies the flooding behaviour by setting a probability of retransmission to a value less than one. There had since been further improvement in such schemes by intelligently choosing the probability of retransmission. Second, introduces a random delay before forwarding, and inhibits forwarding by the potential relay if a certain fixed number of retransmissions had been overheard by that relay. Third, a greedy distance-based scheme,

where after introducing a random delay, the relay retransmits only if the distance between the relay and the closest overheard retransmitter is above a threshold. This scheme is similar to the scheme proposed by Briesemeister *et al.* [84]. Fourth, a location-based scheme where potential relays calculate circle-circle intersections of all retransmissions heard and retransmit only if the potential relay can provide at least some threshold amount of additional coverage. This is similar to the technique presented in this thesis. Finally, cluster-based schemes, where only a subset of stations ("gateway") are permitted to retransmit, and using any of the previous four techniques to determine whether to relay the message. These schemes require some sort of coordination amongst stations in electing gateways and determining cluster membership, and are implemented in works such as [85–87]. In addition to Ni *et al.*'s techniques, flooding in geocast can also be controlled by restricting the set of potential relays.

Restricting potential relays

Perhaps one of the earliest flooding-based technique for geocast that requires no *a priori* knowledge of the network or prior collaboration was presented by Ko and Vaidya [88]. In their work, a source station would address a packet to a defined geographical area, and in addition, define a "forwarding zone" within which all stations would flood the packet to all its neighbours. However, without *a priori* knowledge of the connectivity between stations in the network, the source station would only prescribe the area between the source and the target area as the forwarding zone, potentially missing the only available forwarding station that lie outside the forwarding zone.

Boban *et al.* proposed an alternative method of restricting membership into the set of relays — stations were chosen based on their physical characteristics. This scheme exploits the station characteristics unique to VANETs, choosing tall vehicles as relays only [89]. Based on physical measurements [60], Boban *et al.* identified that tall vehicles are less affected by shadowing than short vehicles,

making tall vehicles similar to semi-trailers good candidates for relays.

Gossipping protocols

Another method of controlling channel load caused by retransmission is to make retransmissions probabilistic. These group of strategies is often known as "gossipping protocols". Haas *et al.* presented one of the first works that made use of the theoretical property that a message would be eventually sent to all stations if each station retransmit with some large enough probability $p < 1$ in "nice" graphs [90, 91]. They then used it to reduce the interference introduced by wireless flooding. Haas *et al.* implemented gossiping on top of AODV, and claimed the technique can reduce message passing by up to 35% compared to flooding. On the other hand, Chandra *et al.* use the technique over MAODV to introduce redundancy into the system, improving error recovery in multicasts [92]. Luo *et al.* implemented the technique over DSR and showed similar resilience to packet losses in multicasts [93].

Recent development, especially for gossipping in VANETs, attempts to improve the adjustment of forwarding probability with minimal prior knowledge of network topologies. Birman *et al.* noted the bimodal behaviour of gossipping protocols — a message will either be completely broadcasted or dies out — depends strongly on the chosen forwarding probability [94]. Furthermore, the critical forwarding probability is highly dependent on the network topology. In order to help resolve this problem, Kyasanur *et al.* presented Smart Gossip [95], which chooses different forwarding probability for each station for each packet with the probability being inversely proportional to the number of duplicates received for the last packet. Parent-sibling-child relationships needed to be maintained to facilitate forwarding.

Kyansanur *et al.*'s work uses one-hop information to determine parent-sibling-child relationships. The selection of forwarding probability can be improved by considering two-hop information. Two-hop topology information can be used

by a relay to determine the probability that its neighbour had already received the packet, and thereby moderating the retransmission probability. Bako *et al.* devised Advanced Adaptive Gossipping (AAG) based on this concept, and showed an improvement in packet reception ratio [96].

Another method to improve on Kyansanur *et al.*'s work is to permit neighbourhood information to change over time. Bako *et al.* [97] changed the neighbourhood discovery process in Smart Gossip [95] into a continuous process, thereby permitting the Smart Gossip protocol be used in MANETs and VANETs [97]. Furthermore, by focusing the work on VANETs, they exploited the property that vehicles typically only move along roads, thus the neighbourhood information can be derived easily from positioning data. Given the importance of positioning information in their algorithm, they named their algorithm "Position-based Gossipping" (PbG). Road intersections present a challenge to PbG because vehicles are no longer confined to mostly 1-dimensional lines, making the derivation of neighbourhood relationships more difficult. Bako *et al.* resolved this problem by adding a second neighbourhood table, enabling PbG to operate on crossroads [98]. In Speed Adaptive Probabilistic Flooding algorithm [99], Mylonas *et al.* exploited the relationship between vehicle speed and vehicle density using traditional traffic flow theory. This allowed the algorithm to estimate neighbourhood information based on locally obtainable information, making the algorithm less dependent on reliable communication between vehicles than PbG or Smart Gossip.

Finally, Bako *et al.* investigated the interference introduced by gossipping protocols. They remarked that as network densities increase, one may safely lower the forwarding probability correspondingly, thereby reducing the network load without compromising on packet reception ratio [100]. Using empirical data from simulations, Bako *et al.* devised a logistic function to moderate the forwarding probability of PbG. Unfortunately, determination of station density in VANETs is not straightforward, especially in the presence of constantly changing pseudonyms as introduced by various privacy measures.

Distance based strategies

Distance based strategies use the distance of a potential relay from some fixed point, and may or may not require this information to be communicated explicitly.

Briesemeister *et al.* extended Ko and Vaidya's work on forwarding zones [88] by concentrating on the scenarios where the source is within the intended target area, removing the need to explicitly define a forwarding zone. In addition, their algorithm implicitly selects a narrow band of stations as relays, and is now considered the classic greedy forwarding algorithm for geocast in highly mobile wireless ad hoc networks such as VANETs. Similar to the work presented in this thesis, their algorithm [84] involves each station making their retransmission decisions independently based on inputs received at the station without prior collaboration. Their retransmission algorithm operates on top of the standard CSMA/CA as specified in IEEE 802.11 and when a packet is received, the station applies a delay before submitting the packet for retransmission. The delay value is calculated using Equation 3.1 [84, Equation (1)] where d is the distance the potential relay is from the original sender. This delay function is inversely proportional to the distance, thus closer stations wait longer. If another copy of the packet is received during the waiting period, the potential relay cancels the retransmission. This behaviour of the algorithm effectively ensures that only the stations at the border of the physical transmission range participate in retransmissions, minimising the hop count of multi hop transmissions. This algorithm also incorporates a hop count field in the header to prevent the packet being forwarded indefinitely. Unfortunately, while this technique does not require *a priori* knowledge of the network, the design of the delay equation will necessarily cause packets to be forwarded to stations outside the intended transmission area ("$MaxRange$" in the paper).

$$WT(d) = -\frac{MaxWT}{Range} \cdot \hat{d} + MaxWT \tag{3.1}$$

$$\hat{d} = \min d, Range$$

where $MaxWT$: maximum waiting time

$Range$: transmission range

Briesemeister and Hommel then improved their work by introducing the concept of dynamically adjusting the intended transmission range of a packet, thereby potentially reducing the size of intended coverage area, hence channel contention [101]. Using emergency braking as an example, they derived Equation 3.2 to be the required transmission range. This concept is effective in identifying the targets of the message more precisely, and improves reception by reducing channel contention. This technique is orthogonal to forwarding algorithms, so can be applied to most similar algorithms.

$$dist_{brake}(v) = v \cdot \Delta t_{reaction} + \frac{v^2}{2 \cdot b_{max}} \tag{3.2}$$

where $\Delta t_{reaction}$: reaction time of driver = 1 s

b_{max}: maximum deceleration = 4.4 m/s^2

Wisitpongphan *et al.* presented a distance based gossipping scheme [102]. Three variants of their scheme were proposed — p-persistent, slotted 1-persistent and slotted p-persistent techniques. The p-persistent scheme is a gossipping protocol, with the forwarding probability p being a linearly increasing function of distance from source. The slotted 1-persistent scheme is essentially Briesemeister's algorithm [84], except the delay value is quantised. Finally, the slotted p-persistent method is initially the same as the slotted 1-persistent scheme, but if a retransmission was overheard during the delay, the relay would forward with

probability p (same calculation as the p-persistent scheme) instead of discarding it. All three variants favour the furthest station for retransmission. The slotted timeline helps synchronise the stations such that retransmissions can be more reliably detected.

For unicast routing (not geocast), it is possible for greedy techniques like Briesmeister *et al.*'s [84] to fail as there may be no station closer to the destination than the current station, and the packet needs to be routed around some obstacles. Perimeter routing [103–105] and face routing [106–109] techniques were developed to address this issue, guaranteeing loop-free paths. These techniques are not relevant for geocasting, and will not be discussed further.

Location based strategies

Briesemeister *et al.*'s algorithm [84] can also be improved by further restricting the potential relays. In M-GeRaF [110], Odorizzi and Mazzini adapted GeRaF [111], restricting potential relays by also considering the direction of the message propagation. M-GeRaF excludes stations that induce a large change in direction as candidates. For example, if a packet was last forwarded towards the east, the packet will not be forwarded by a station that would make "progress" towards the west. M-GeRaF extends the original GeRaF by allowing multiple sink stations. This improvement is able to reduce interference by reducing retransmissions that make minimal progress, but will still introduce interference outside the target area due to the underlying progress metric.

Ni *et al.*'s example of location based strategy [82, 83] was for each station to keep track of all overheard retransmissions. Assuming that the potential relay knows the centre of the overheard transmission (*e.g.* by incorporating relay position in a header field), and estimating the range of that transmission (*e.g.* by using receive power), the technique forms a set of overlapping circles. Ni *et al.*'s algorithm then computes the additional coverage provided by the potential relay (area of the circle centred at the relay, less overlaps with any other circles). Poten-

tial relays are prevented from forwarding the message if the calculated additional coverage is below some threshold.

Ni *et al.* noted that the computation of circle-circle intersections with many overlapping circles is expensive. They suggested the use of either a grid filling approximation or by using convex polygons to help determine whether to re-broadcast. The method using convex polygon involves determining whether the potential relay lies within the convex hull of the polygon containing all the centre points. If the relay is outside the convex hull, then it is permitted to transmit, otherwise, retransmission is suppressed. Ni *et al.* calculated that geometrically, this technique can cause up to 22% of stations not to receive the packet.

Ni *et al.*'s original technique is very expensive computationally, and the convex hull approximation is still quite expensive, even though there are $O(n)$ algorithms that can compute the convex hull. The convex hull approximation requires all potential relays to track all retransmissions, therefore has a higher memory requirements as well. Furthermore, this technique also does not encourage retransmission by preferable stations, it just suppress the undesirable ones, therefore can still introduce significant redundancies. Being a broadcast algorithm (not geocast), this technique forwards irrelevant packets outside the intended coverage area.

Urban Multi-Hop Broadcast (UMB) [112] and its extension Ad-hoc Multi-Hop Broadcast (AMB) [113] are two of the most well-cited protocols that uses the location-based technique for controlling contention. UMB operates primarily as a distance-based scheme, but adds the requirements for repeaters to be installed at intersections to help overcome building shadows. In AMB, Korkmaz *et al.* removed the need for repeaters by incorporating an algorithm to identify relay vehicles near intersections to replace the repeaters. Both UMB and AMB borrowed the RTS/CTS scheme in IEEE 802.11, and introduced a similar RTB/CTB (Request-to-broadcast/Clear-to-broadcast) handshake. RTB/CTB operates similar to RTS/CTS, except the RTB is untargeted and CTB is generated by all

potential relays in distance order with furthest relay transmits first. CTB inhibits all other transmissions including CTB from closer relays. Once RTB/CTB handshake has been completed, the source begins broadcast, and the relay who sent the CTB is responsible for forwarding it. RTB/CTB helps suppress unnecessary forwarding and reduces collisions in the same way as the RTS/CTS virtual carrier sense. This method of selecting relay stations was also used by Taha and Hasan [114] as well as Fasolo *et al.* in Smart Broadcast [115]. Barradi *et al.*'s Highway Multihop Broadcast (HMB) [116] uses a variant of RTB/CTB, allowing the RTB/CTB to be routed to cover the target area, and also use these RTB/CTB to introduce QoS access control. HMB suppresses the broadcasting of lower priority messages by stations receiving the RTB/CTB. Unfortunately RTB/CTB-based broadcasts require an explicit handshake, which may delay the delivery of time-sensitive data.

Máté and Vida combined the idea of location-based forwarding with gossipping technique to form the Localized Urban Dissemination scheme (LUD) [117]. LUD contains elements similar to the intersection determination in AMB such that only stations near an intersection may participate in the probabilistic forwarding. LUD also assumes the availability of a database containing traffic statistics, especially turning probabilities at each intersection, and uses information in that database to assist in assigning forwarding probabilities. This work was then extended by also adopting a counter-based retransmission inhibition [118].

Sung and Lee's Light Weight Reliable Broadcast Message Delivery (LW-RBMD) [119] combines the intersection relay priority found in AMB [113], and relies on a traffic database similar to LUD [117]. Their technique uses two separate timers to determine retransmit priority — one for the distance-based implicit ACK, the other for intersection priority. While their investigation suggested an improvement in delivery efficiency (in terms of number of packets received per broadcast message), their scheme is more complicated to implement than other schemes of this type.

DV-CAST [120] is another location-based technique. In DV-CAST, a station determines its one-hop neighbourhood topology by observing periodic beacon messages. It then uses this information to switch between three behaviours — well connected (has potential relay in direction of propagation), sparsely connected (no relay in direction of propagation, but has neighbours in the opposite direction) and disconnected neighbourhoods. When a potential relay is "well connected", it uses a slotted 1-persistence algorithm (forward after a delay unless suppressed by overhearing another retransmission). In the sparsely connected mode, it rebroadcasts immediately, and depending on configuration, might wait for a further rebroadcast from new opposite direction neighbours or simply return to the idle state. When a station is disconnected, it would cache the packet until either the packet timer expires, or retransmit when a new station arrives. DV-CAST demonstrated the use of locally-obtainable neighbourhood information in order to adjust the potential relay's behaviour, and may be a useful extension to the technique presented in this thesis.

The interference-aware retransmission algorithm in this thesis is a location based technique, using the time-delay prioritisation technique from Briesmeister *et al.*'s work [84], and borrows the circle-circle intersection concept from Ni *et al.*'s work [82, 83]. Unlike previous techniques, the technique presented in this thesis does not require tracking of all retransmissions, and is therefore more resilient to errors in tracking other stations. It encourages retransmission by desirable stations, while at the same time suppresses undesirable stations. The algorithm only remembers the SINR of the strongest copy of each packet in order to further suppress unnecessary retransmissions.

Cluster based strategies

Cluster-based techniques try to reduce communication cost by restricting long-distance communications to a certain subset of stations, each of those are responsible for the stations in their cluster. Liao *et al.* [85] presented GeoGRID, a

cluster-based routing technique whereby stations implicitly belongs to the cluster defined by geographical grid cells. Liao *et al.* defined an election algorithm to select cluster heads. This technique requires the use of fixed-sized grids, thus cluster membership may become highly skewed.

Further improvements on this type of systems include improving communications within the cluster and between clusters, as well as improving the way stations are partitioned into the clusters. For example, Jain *et al.* [121] proposed the use of a Voronoi diagram to assist with partitioning, and later Stojmenovic *et al.* [86] used Voronoi diagrams followed by convex hull heuristics to help inter-cluster communications. Hoang and Motani investigated additional aggregation techniques that can be employed as part of the cluster [122]. Santos *et al.* presented a variant of the cluster-based algorithm whereby messages are passed between clusters through "gateway" stations that belong to multiple clusters [123].

Similar to predetermined routes methods, cluster-based techniques are not primarily designed for ensuring reliability, but rather reducing redundancy. Cluster heads have the responsibility for message delivery within the cluster, but packet losses to cluster heads would be detrimental to the communications system. Mauve *et al.* remarked that fault resilience can be built into the communication system by increasing the overlap between clusters such that stations overhear communications of multiple clusters [87]. However, this would greatly negate the benefits provided by such cluster-based systems.

Aggregation, Concatenation and Piggybacking

A final class of methods for reducing interference in cooperative forwarding is to completely alter the packets being forwarded whilst maintaining the information they contain. There are two main ways to achieve this goal — piggybacking and aggregating. In the piggybacking scheme, packet retransmissions are attached to other packets from the relay, potentially greatly reducing the overall contention by reducing the contention events. Jiang *et al.* first proposed the use of piggy-

backed ACK/NACK responses for single-hop broadcasts, which can be use to trigger retransmissions as needed [124]. Yang *et al.* presented and evaluated a piggybacking retransmission method [125]. In their technique, potential packets to be relayed are ranked based on a certain metric (*i.e.* earliest deadline first and furthest distance first, but is adaptable to any given metric). On expiry of some timer, a fixed number of packets are either concatenated together or piggybacked to packets from the relay based on the metric used and queued for retransmission. They found by simulation that such a piggybacking scheme can improve reception ratio and reduce collision. These relaying techniques operate orthogonally to the actual retransmission algorithm, and can be used to augment algorithms including the one presented in this thesis in order to further reduce channel contention.

Unlike piggybacking and concatenation, the concept of aggregation is to process the message contained within the packet in the application layer, then forward a summary of that data instead of blindly forwarding the packet at the network layer. Wischoff *et al.* presented a system whereby application data are subdivided according to road segments, with results aggregated [126]. Unfortunately this system suffers from poor scalability [127], and is not easily adaptable to different application requirements [128]. Improvements to aggregation systems have been published on enhancing scalability through aggregation hierarchy for comfort applications [129, 130]. Since data aggregation is about presenting a summary of the underlying data, it is not suitable for many safety applications.

Delay-tolerant networks (DTN)

For completeness, it is also noted that there is a body of work related to "delay-tolerant networks" (DTN) [131, 132]. Delay tolerant networks deal with the dissemination of messages through physical movement of relays in order to minimise data transmission. These techniques are useful for information that are not delay sensitive and are relevant to either a large area, or in a distant area. In DTNs,

relays are usually selected based on their mobility characteristics (*e.g.* speed [132], current location [133], direction [133], *etc.*) in addition to their propagation desirability (*e.g.* height). Due to the delays involved in these techniques, they are typically unsuitable for safety messages, and will not be further discussed in this thesis.

3.2 Adapting to channel contention

Regardless of the retransmission scheme used to enhance packet reception, the act of retransmission introduces extra load onto the radio channel. When the lower-level communication protocol uses a contention-based channel access scheme, the extra load increases the contention on the radio channel, which can suffer degradation in packet reception due to timeouts and/or collisions. Whilst investigating the various aspects of the design of DSRC, Jiang *et al.* had identified that the principal factor affecting CSMA/CA broadcasts is "communication density" (the "number of carrier-sensible events per unit area and unit time", *i.e.* the product of message rate, station density, and range) [124]. For this reason, the ability to determine and adapt to the current "communication density" (*a.k.a.* channel contention) can greatly enhance the performance of the network.

Similar to many consumer wireless devices, DSRC VANETs use contention-based channel access schemes to coordinate access to the wireless channel. The IEEE 802.11 standard specifies the use of the Enhanced Distributed Channel Access Function (EDCAF), which is backwards compatible with the older Distributed Coordination Function (DCF) for transmissions that do not have QoS requirements. Both DCF and EDCAF are implementations of the CSMA/CA scheme for contention-based channel access. While these contention-based schemes benefit from lower control overhead, their channel utilisation tend to be sub-optimal when the contention is low, and their throughput can quickly degrade as the channel becomes congested.

CSMA/CA-based schemes, including the IEEE 802.11 EDCAF, use a variable-

length contention window to in order to adapt to channel contention, thereby balancing channel utilisation and throughput. The correct sensing of current channel contention is therefore important for the optimal adjustment of the contention window size. Reinders *et al.* had identified that under the default IEEE 802.11p broadcasting contention window (16 slots) and a beacon generation rate of 10 Hz, single-hop packet reception ratio falls below the 99% requirement due to contention when there are only 20 vehicles per km per lane on a 4-lane highway [134]. Reception falls below 50% at around 120 vehicles per km per lane on the same 4-lane highway.

This thesis contains works on improving the resilience of channel contention estimation technique. Even though many existing works had concentrated on the analysis and improvement in channel sensing and adaptation, most of these works are developed for unicast transmissions due to the dominance of this transmission mode in consumer wireless devices. Channel sensing and adaptation on broadcasts are much less studied. In VANETs, many applications use broadcasts or geocasts instead of unicast. These transmission modes present unique challenges that cannot be addressed using existing sensing and adaptation techniques.

In this section, the large body of work describing methods to adapt the DCF in response to changing channel contention is surveyed. First the IEEE 802.11 standard itself has a mean to adjust DCF parameters to adapt to varying contention, but it only works for unicast frames. Second, a range of preemptive contention window adjustment techniques designed to improve the performance of the IEEE 802.11 DCF is discussed. Here, some of these approaches simply assume that the level of channel contention is known [135], while others also describe ways to sense channel contention. Third, the current techniques of measuring channel contention, including observing frame collisions [136–139], and the channel utilisation [140, 141] are compared and contrasted.

3.2.1 The Standard IEEE 802.11 DCF/EDCAF

Even though the IEEE 802.11 standard does not provide a direct metric for channel contention, the DCF has mechanisms to adapt its behaviour based on channel contention, hence one can deduce information on channel contention by observing system parameters. The IEEE 802.11 DCF adapts to channel contention by moderating the CW size — a large CW reduces the probability that more than one station will transmit concurrently (hence frame collision), whereas a small CW improves throughput for the sending station. A binary exponential backoff mechanism is used to expand the size of CW when the channel is perceived to have high contention. At the completion of a frame's transmission attempt (either successfully or have reached the maximum number of retries), the size of the CW is reset, allowing the station to improve throughput should the channel become less congested. Since DCF adapts the CW based on perceived contention, the CW size at the completion of a transmission attempt can provide an indirect measure of the channel contention.

Unfortunately, as mentioned in Section 2.5, the exponential backoff mechanism is not triggered for broadcast frames, and therefore the CW size is not a suitable measure of channel contention for these broadcast communications. The assumption that frame losses are caused only by frame collisions is also not always true as it ignores frame losses due to distance (path loss), obstacles (shadowing) [55, 56, 142] and fast fading [53, 143], potentially overestimating channel contention. It also ignores frame capture [144, 145] — the ability for the physical layer to successfully decode a strong enough signal even if that frame collided with a weaker signal (typically due to the near-far effect) — and potentially underestimate the channel contention.

3.2.2 Slot Utilization

To improve the performance of IEEE 802.11 networks, many works have been presented to enable dynamic, pre-emptive setting of the CW based on channel

contention. Amongst the first works addressing this is Bononi *et al.*'s [140] work on the "Distributed Contention Control (DCC)" mechanism, which features the metric "Slot Utilization" (SU) as a mean of determining channel contention. Bononi *et al.* initially defined this metric as $Slot\_Utilization = \frac{Num\_Busy\_Slots}{Init\_Backoff}$. This metric is designed to be computed when a station has a frame to send. The frame follows the normal DCF process and is assigned an initial backoff from within the CW. As the station counts down its backoff period, it observes the channel for other transmission attempts during that period. It relies on the observation that as the number of contending stations increases, the more backoff slots will be occupied. Bononi *et al.* claimed the Slot Utilization metric to have the following properties:

1. values are within $[0, 1] \in \mathbb{R}$ — 0 indicates no slots were occupied, and 1 indicates transmission attempts were observed on all slots; and

2. intermediate values within $[0, 1]$ should be proportional to the contention level.

The Slot Utilization metric is useful in providing a contention measure that can be used to compare different contention levels. Since this metric only looks at transmission attempts, it does not consider the success or failure of that attempt and is therefore more resilient to physical layer channel conditions. This metric was originally designed to complement link layer congestion control measures, therefore the authors specified that it be obtained based on the selected backoff value. For continuous channel monitoring, this may cause a very slow update rate (it would update only if the station have something to send), and may have large fluctuations as the initial backoff value is randomly chosen by the station. Averaging this measurement over a larger measurement window may improve this usefulness of this metric for continuous channel monitoring.

Unfortunately, the assumption of linearity between channel contention and the proportion of busy slots (property 2) does not hold. As the channel contention

increases, collision probability increases, therefore the mean proportion of busy slots is asymptotic to 1, and is therefore not proportional to the contention level as asserted. In fact, as will be shown in Chapter 5, the relationship between the number of contending stations and the mean proportion of busy slot is actually quite complex and definitely non-linear.

Bononi *et al.* further extended their concept of Slot Utilization, developing an "Asymptotic Optimal Backoff" (AOB) mechanism [146]. AOB uses the SU metric and extends on the p-persistent IEEE 802.11 MAC [147]. AOB aims to set the parameter p in the p-persistent MAC dynamically by tracking slot utilization, making an assumption on the packet size (affecting the parameter q). The mechanism is asymptotically optimal in the number of stations (M) multiplied by determined optimal parameter (p_{min}). They have determined that AOB estimation of $M \cdot p_{min}$ improves as message length (q) increases. They have also investigated the transient behaviour of AOB, and found that AOB copes well when the number of concurrent stations changes sharply, contrasting with the standard MAC's inability to cope.

3.2.3 Methods based on theoretical analysis

In parallel with the work on heuristic based MAC layer improvement technique such as [140], others have approached the problem of improving the DCF through theoretical analysis. Bianchi *et al.*'s discrete-time Markov Chain model of the IEEE 802.11 DCF [148, 149] (Figure 3.1) is one of the most well known models, and is still being used and extended for other applications. Bianchi *et al.*'s model discretises time into slots, analogous to the slots in the DCF, and has layers of longer chains to model the DCF recovery process (binary exponential backoff). The model assumes saturated stations. There are further works extending this model by relaxing the saturation assumption [150–152] (Figures 3.2, 3.3), relaxing the "perfect channel" assumption by considering frame loss due to noise (modelled as a probability) [153], as well as incorporating extra queues to model

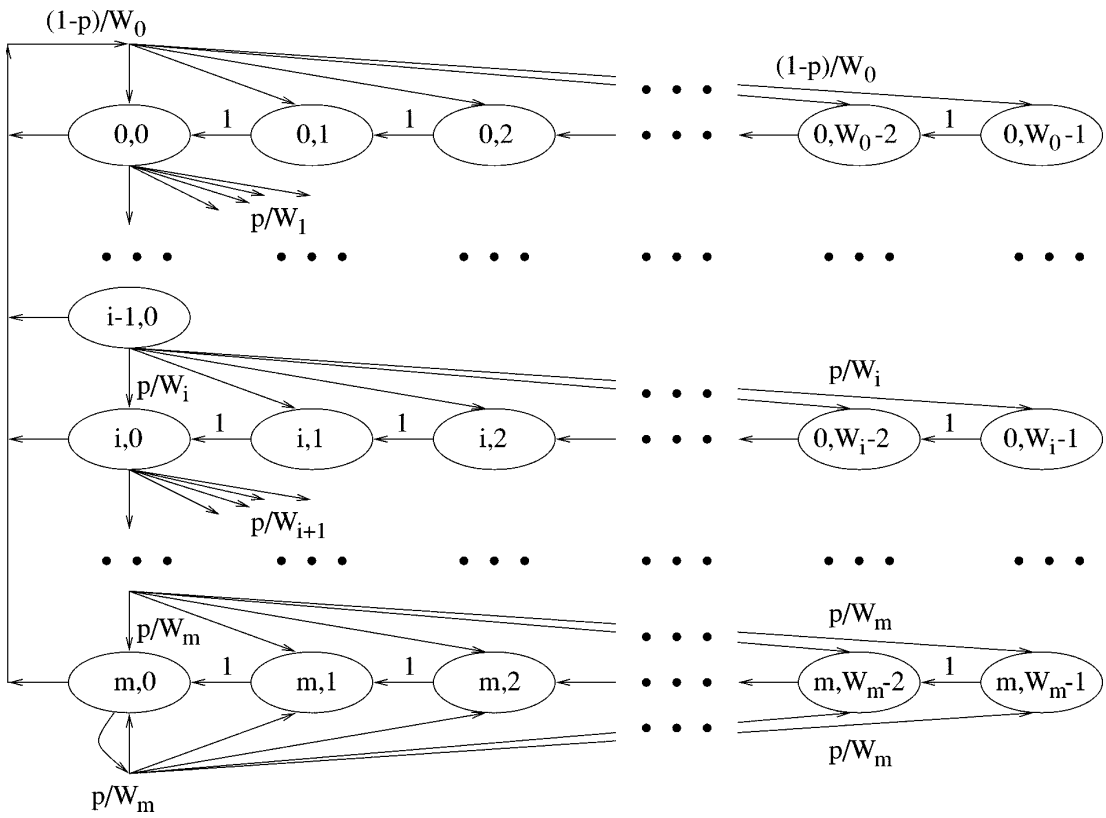


Figure 3.1: The Bianchi Model. [148, Figure 4] — The top row is the initial contention window. Subsequent rows represent retry attempts. The modelled station fails with probability p, after which the station enter the next level of retry attempt. The Markov model does not model the maximum retry limit of a station.

EDCAF [152]. In addition, investigations on the veracity of the assumptions used in these models had recently validated the models against ns-2 simulation and test bed data, finding the model to be reasonable for both saturated and unsaturated stations with no buffers [154–156].

Unlike Bianchi *et al.*, Ma and Chen modelled IEEE 802.11 broadcasts, which are expected in VANETs [52]. In their work, they constructed two Markov chain model of message processes, one for "emergent services" and the other for "routine services" (Figure 3.4). In order to accommodate for the unsaturated nature of these messages, Ma and Chen used a separate Poisson processes to model packet arrivals. The authors then solved the system as an M/G/1 queue, deriving various metrics of the system [157].

Ma and Chen's model uses Poisson processes to model packet arrivals, which

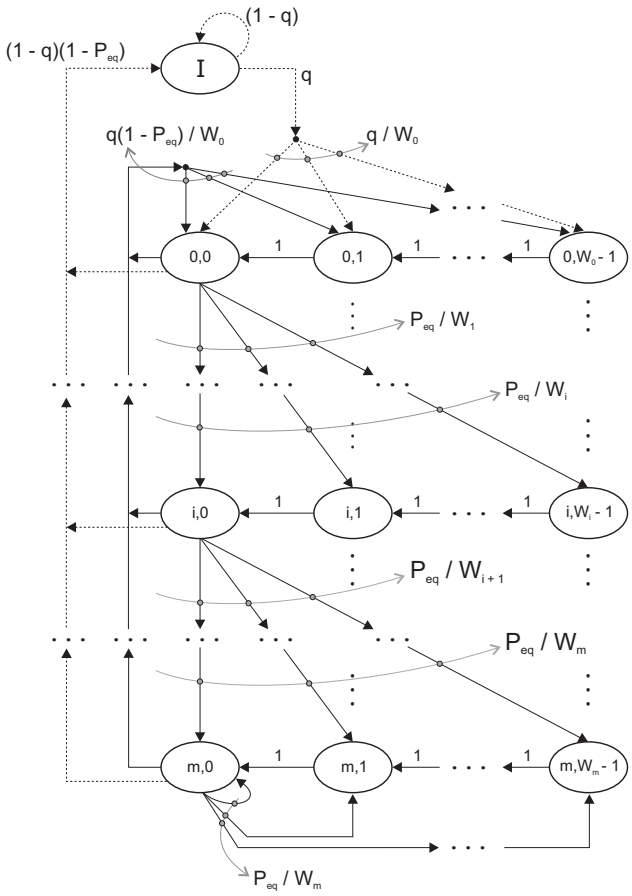


Figure 3.2: Extension to the Bianchi model to account for unsaturated stations [150, Figure 1] — An extra waiting state is added above the standard Bianchi model [148]. This state represents the time where the station does not have anything to send. When a packet arrives (with probability q), the station commences standard MAC backoff and then follows the Bianchi model. This model does not account for the allowance for stations to decrement the backoff counter when has nothing to send.

is not suitable for modelling periodic transmissions such as beacons. In Bastani *et al.*'s model [51], they introduced a fixed length chain prior to the branching behaviour of the DCF backoff process (Figure 3.5). This fixed length chain introduces a fixed delay, giving a more accurate model of beacon messages. Their model is also discretised into slot times (hence assumes message transmission takes integer multiple of slots), but at each slot time, the DCF backoff state does not advance if the channel is busy. For an extremely congested channel, a station may remain in the backoff stages for a very long time before a new beacon is generated. For this reason, the model is still unable to fully capture the periodic nature of beacons.

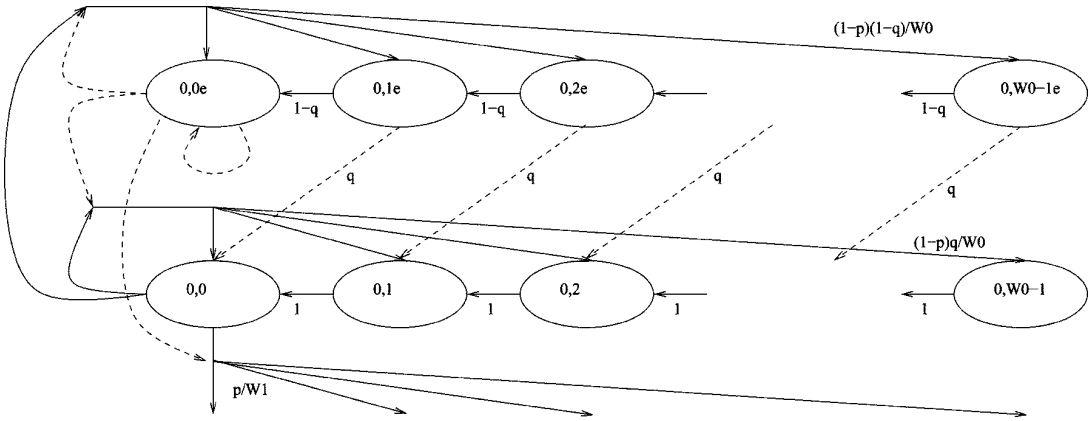


Figure 3.3: Malone *et al.*'s extension to the Bianchi model to account for unsaturated stations [152, Figure 1] — The top row (states ending in "e") is the addition to the Bianchi model [148]. This top row represents the time where the station does not have anything to transmit, but is still counting down the backoff timer. When the MAC receives a frame to be transmitted, it jumps to the next lower state in the standard initial backoff state and follows the usual process as per the Bianchi Model. At state (0,0e), possible actions are: 1. still have no frame to transmit (loop back to self with probability $(1-q)$); 2. unsuccessfully transmits (go to branching below row 2 with probability pq); 3. successfully transmits and has no other frame ready (back to top branching with probability $(1-p)(1-q)$); and 4. Successfully transmits and has another frame ready (to second branching with probability $(1-p)q$).

A recent paper by van Eenennaam *et al.* [158] described another method to analyse VANET beacons based on Markov chains, and is similar to the work in this thesis. Similar to Ma and Chen's model [52], they have defined unsaturated stations so as to allow random message arrival from a Poisson process. The resultant Markov chain from their analysis (Figure 3.6) is not too dissimilar to Malone *et al.*'s [152] except for the lack of the reattempt layers. Differences between numeric and simulated results for their unsaturated test highlight the difficulty in generalising such MAC layer analysis to unsaturated stations.

Based on these models, the transmission probability of each station can be evaluated, and more importantly, expressions for the collision probability as a function of contending stations can be derived.

These works therefore suggest that one can use observed collision probability as a measure of channel contention. Following from these works, Bianchi and

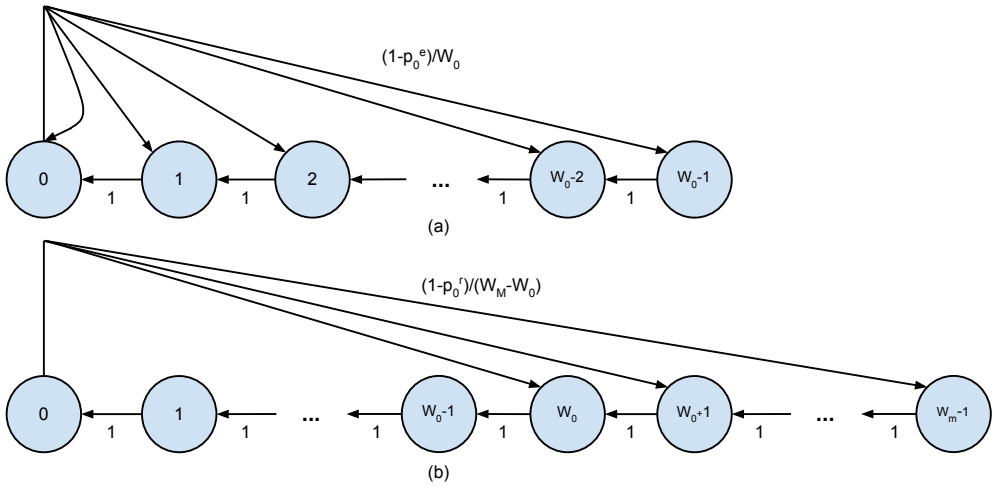


Figure 3.4: Ma and Chen's Markov model of broadcast messages (two QoS classes) [52, Figure 3] — Exponential backoff is not modelled as they do not happen in IEEE 802.11 broadcasts.

Tinnirello presented a method to estimate the number of contending stations based on observing the conditional collision probability [136]. In this work, the authors rearranged the collision probability function to yield a function that maps collision probability to the number of contending stations. The authors then demonstrated two implementations of this function and showed by simulation that both techniques are effective in obtaining and tracking an estimate of the number of contending stations even when channel contention is varied in the scenario.

Both channel contention sensing methods presented by Bianchi and Tinnirello [136] operate by observing the radio channel. They estimate the probability of collisions by assuming the observing station is also about to transmit a frame, hence the "collision probability" can be inferred by simply observing the channel busy status. These two methods differ by the way the measurements are aggregated — one uses an Auto Regressive Moving Average (ARMA) filter, whereas the other uses an Extended Kalman Filter (EKF). Simulations on the ARMA filter implementation showed that as the number of contending stations increases, the noise from the filter output increases. This is as expected due to the reducing slope of the formula. To mitigate this noise, the authors implemented the technique using an EKF. An Extended Kalman Filter enables runtime ad-

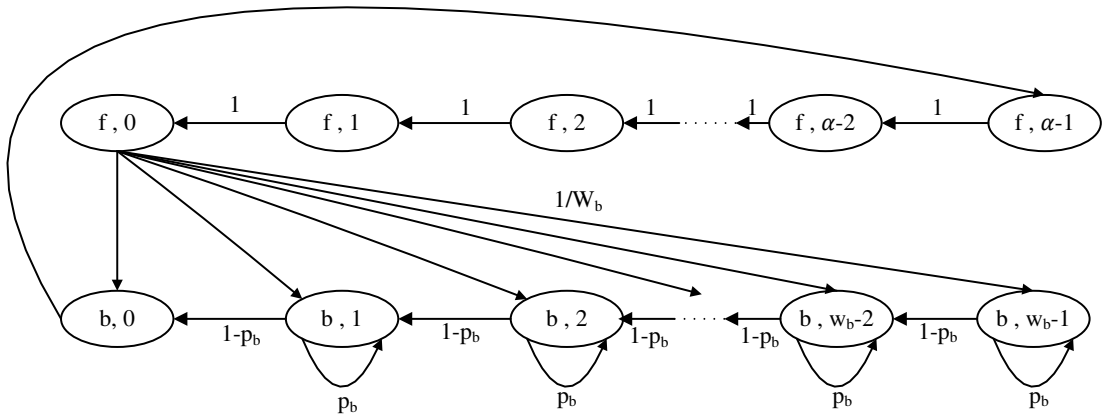


Figure 3.5: Bastani *et al.*'s Markov model of safety messages broadcasts (beacons) [51, Figure 1] — Unlike previous models, stations under this model do not always count down the backoff states depending on channel conditions. This model requires close coupling between the channel states and the model parameter, requiring a more complex method to solve the system. The extra non-branching chain on top represents waiting states that is used to model fixed delay-periods for beacons. This still cannot fully capture the periodic behaviour because the delays encountered in the bottom chain are variable, making the period of the broadcast non-deterministic.

justment of the tolerable noise, something an ARMA filter cannot do. Finally, the authors presented a change detection routine. When changes in the amount of contention are not detected, the routine increases the filter memory, thus uses all new updates to improve the previous estimate. When a change is detected, the filter quickly reduces its filter memory, allowing the filter to quickly update to the new contention level. The authors also tested both implementations on stations with unsaturated traffic with packet arrival at the station following a Poisson process, and showed that both implementations are able to provide an estimate of the "average" number of contending stations.

Some more recent work have extended Bianchi and Tinnirello's techniques by substituting alternative filters for the ARMA and the EKF. Toledo, Vercauteren and Wang investigated the use of batched and sequential Bayesian estimators combined with maximum a posteriori (MAP) [137], sequential Monte Carlo techniques, and Viterbi algorithm [138]. Kim, Serpedin and Shin also investigated the use of various particle filtering techniques for this task [139].

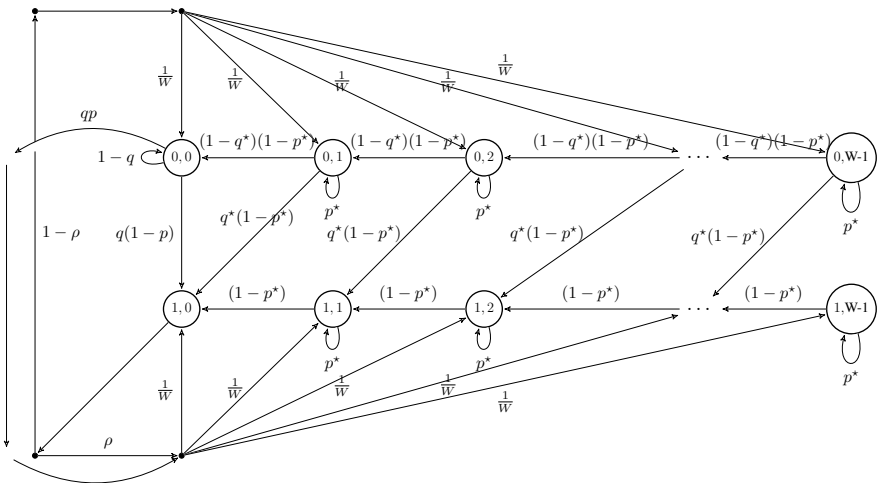


Figure 3.6: van Eenennaam's Markov Markov model of VANET beacon messages [158, Figure 3] — This is similar to Malone *et al.*'s model [152] without the retry attempts.

This group of techniques had demonstrated their accuracy and effectiveness in tracking the contention level, but relies heavily on collision probability. The metric (number of contending stations) can be easily interpreted, and can provide intuitive feedback to upper layers in order to moderate their offered load. As described previously, techniques based on Bianchi and Tinnirello [136] observes whether the channel is busy as a proxy for packet collision. This assumes that the observer station is going to transmit a frame regardless of the channel condition, hence the observer station will cause a collision if the channel is busy. These imagined transmissions make the observer station oversaturated as the station does not follow the CSMA/CA backoff rule and would send a packet at every slot. This therefore artificially inflates the amount of channel contention sensed, as confirmed in Chapter 5.

Furthermore, Bianchi *et al.* had also identified in an experimental assessment that commercially available network cards have a tendency of not conforming to IEEE 802.11 backoff behaviour [45]. This non-conformance is likely to impact on the usability of most contention-estimation techniques (including the one presented in this thesis) in real world applications. Conformance tests on backoff behaviour should therefore be standardised, and is highly recommended prior to any real world deployment of DSRC hardware.

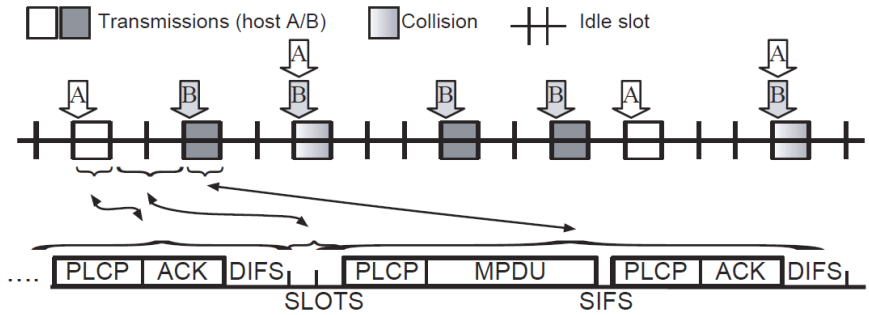


Figure 3.7: Two stations contending for the same channel [141, Figure 2] — Idle Sense counts the "SLOTS" as shown in the figure, and adapts the CW size in order to balance channel contention and throughput.

3.2.4 Idle slots

Heusse *et al.* presented a variant of IEEE 802.11 DCF improvements technique [141] — instead of measuring and tracking the number of contending stations, it is based on adjusting the CW continuously to achieve an optimal level of contention. Extending from the Bianchi model [148], Heusse *et al.* derived an expression linking throughput to the number of idle slots between transmissions (Figure 3.7), and presented an algorithm to maintain the CW size so as to keep the measured number of idle slots between transmissions close to the optimal value. While this work does not directly address the issue of measuring channel contention, it suggested that given a fixed CW (which is expected for broadcasts and geocasts), the number of idle slots between transmissions is an indicator of channel contention. Compared to Bianchi and Tinnirello's technique [136], this method of measuring contention does not rely on observing collisions, and is therefore both immune to the frame capture effect, and does not make assumptions on the observing station. Compared to Slot Utilization [140], this method does not assume a linear relationship between contention and the proportion of busy slots. This thesis extends this idea of observing idle slots counts to measure channel contention.

3.2.5 Differentiating between channel errors and collisions

In addition to the attempts to directly observe/measure the load on the radio channel, there had also been other works that can also infer channel load and/or

improve the estimate of channel load. One class of these techniques attempts to account for frame errors that arise from channel conditions, thereby relaxing the common assumption that frame losses are caused by collisions. Pang *et al.* proposed the use of a new NACK frame to help differentiate the causes [159,160], Ma *et al.* looked at characterising loss statistically [161], while Malone *et al.* proposed an observation-based approach [162]. These techniques complement the vast literature covering techniques that operate at the PHY-layer (*e.g.* [161, 163, 164]), and at higher layers (*e.g.* TCP [165], UDP [166], TFRC [167]). While these loss differentiation algorithms can help provide and improve the feedback on channel contention to the application, many would not work in broadcast scenarios due to the lack of explicit feedback at the lower layers, and therefore provide limited improvement over contention estimates that do not rely on detecting collisions.

As explained in the previous subsections, there are many limitations when trying to adapt and/or use these techniques for measuring channel contention in broadcast situations. To alleviate these limitations, this thesis presents a passive method that relies on overhearing current communications. This technique is similar to the one proposed by Bianchi and Tinnirello [136], but relaxed the requirement for observing "packet collisions", allowing the technique to be potentially more proactive in reducing collisions. This technique also produces a less noisy output and is more resilient in presence of unsaturated stations.

3.3 Accuracy of computer simulations

Computer simulations are used in most research on wireless networking, including most of the existing works outlined in this chapter. Through a survey on papers published in MobiHoc, Kurkowski *et al.* identified that 75.5% of the work on MANETs published during 2000–2005 used computer simulations [168]. However, the accuracy and validity of commonly used network simulator packages have not been well studied despite Johnson [169] having proposed a framework to validate simulation results against real and/or emulated scenarios in a 1999 DARPA study.

Of the few published validation studies concerning wireless networks, the majority focused on the accuracy of the physical layer modelling, which is perhaps the layer that is most abstracted. Liu *et al.* [170] validated the SWAN simulator against testbed results, Rachedi *et al.* [171] compared ns-2, OPNET and QUALNET also against testbed, and both drew conclusions on the accuracy and configuration requirements of the PHY layer models. Pei and Henderson [172] specifically investigated ns-3 IEEE 802.11b PHY model against theory, and concluded that ns-3's IEEE 802.11b PHY model is accurate in line-of-sight scenarios. Ivanov *et al.* [173] found high deviations between simulation, emulation and testbed results from multi-hop wireless network simulation, and similar to previous works, attributed the differences to the PHY layer abstractions.

Attribution of differences to physical effects are not limited to just the PHY layer of the communication stack. In their validation of wireless sensor network in OMNet++ against testbed measurements, Colesanti *et al.* [174] found that OMNet++ is unable to adequately simulate hardware timing "quirks" inherent in the actual testbed stations.

A few other works have validated simulation outputs, either against each other or against testbeds, and drew some conclusions on MAC layer implementations. Cavin *et al.* [175] compared OPNET, ns-2 and GloMoSim and observed highly diverging outcomes amongst the simulators, and conjectured that the deviation stemmed primarily from the variations in abstractions used in their respective PHY models, as well as non standard-compliant implementation of the MAC protocol. Bredel and Bergner [176] validated OMNet++ simulations of IEEE 802.11g unicast communications against testbed results. They found that while the simulation outcomes mostly achieve a good match against testbed results, they observed significant differences in the "inter-transmission" metric when there are more than two saturated stations on the network. (Inter-transmission is the number of packets transmitted by a saturated station during the time taken for another saturated station to transmit two packets. This metric measures

the fairness amongst contending stations.) Bredel and Bergner attributed the difference to the scheduling behaviour of OMNet++ MAC models.

To date, it appears that only one work had been published specifically seeking to validate the MAC layer behaviour of a well known simulator. Baldo *et al.* validated ns-3 simulations against testbed, showing good agreement with testbed observations between the macroscopic behaviour of the MAC layer, as measured from the application layer metrics (*e.g.* throughput, latency, *etc.*) in most test scenarios [177]. For the scenarios where deviations were observed, Baldo *et al.* attributed the differences to both the simulation model and testbed hardware configuration, citing limitations in the version of MadWiFi driver installed on the testbed stations.

Even though many works outlined in Section 3.1 relies on having access to accurate MAC layer statistics, it appears that no existing investigations had been presented validating the MAC layer statistics from well known simulator packages. All the works identified above only investigated the validity of simulation from application layer metrics such as throughput, latency, and inter-transmissions. While these metrics are useful for validating application layer performance results, they do not validate the correctness of the MAC algorithms implemented. It is not possible to be certain about the trustworthiness of evaluations on contention adaptation technique that uses these simulators. In this thesis, it is demonstrated that errors exist in these simulators' MAC implementations and are not easily observable from the application layer statistics. Furthermore, the causes and consequences of such misbehaviours in ns-3.9, are investigated, culminating in a set of workarounds that can be applied in order to correctly use and interpret the simulation outcomes.

3.4 Conclusion

In summary, the work presented in this thesis contributes to the three areas identified above. It describes a channel contention estimation technique that is both

more reactive to changes in channel contention, and is very accurate in predicting non-reception of packets. It addresses the inherent interference generated by cooperative retransmission algorithms, improving the efficiency of retransmissions from relays. A load reactive geocast system is also presented, which allows enhanced packet reception through a channel efficient geocast algorithm, and is capable of adapting to channel conditions. Finally, this thesis highlights the important but hidden problem regarding the accuracy of existing well-known computer simulation packages, and presents workarounds that allow one of these simulators to be used despite its non standard-compliant behaviour.

Overview

This chapter presents and evaluates a geocast algorithm that aims to balance the amount of interference produced with the amount of redundancy required. A control parameter is exposed in this algorithm, enabling control algorithms to be developed that can adjust the tradeoff at run time.

Contributions

- **I developed a metric that ranks wireless stations for relay preferences.** This metric considers both the extra coverage a station provides, and the interference it introduces by retransmitting. In addition, this metric is computed from the potential relay without the need for coordination amongst other stations. It is independent of the actual retransmission algorithm and hence can be used by other algorithms for prioritising relays.

- **I implemented and evaluated a retransmission algorithm utilising the retransmission metric.** The metric was implemented in an ns-3 simulation, together with a delay-based relay selection algorithm. Simulation results showed that the metric is capable of selecting good relay stations in order to cope with high network contention scenarios, with the scalability controlled by a dynamically adjusted parameter.

Publications

- **Tse, Quincy**, "Improving Message Reception in VANETs," in Proceedings of *Mobile Systems PhD Forum, 2009 International Conference on*, Krakow, Poland, Jun 2009.

- **Tse, Quincy** and Landfeldt, Björn, "Interference-Aware Geocasting for VANET," in Proceedings of *World of Wireless, Mobile and Multimedia Networks, 2012 IEEE International Symposium on*, pp. 1-6, San Francisco, CA, USA:IEEE, 25-28 June 2012.

Chapter 4

Interference-Aware Geocasting

4.1 Introduction

This chapter presents a cooperative retransmission algorithm requiring no *a priori* knowledge of the network, yet controls the amount of unnecessary interferences caused. For this algorithm, a retransmission metric that estimates the benefit-to-interference ratio is developed and combined with a delay-based priority scheme in order to select the most appropriate relays. This algorithm improves the reception of periodic beacon messages used in Cooperative Collision Avoidance (CCA) systems, such that relays provide spatial diversity for recovering failed receptions due to distance or shadowing. In this thesis, a station is defined as experiencing interference if its radio interface senses a carrier on the channel, or when it is receiving duplicate and/or irrelevant packets. This interference adds extra load onto the wireless channel, which can potentially reduce packet reception and increase delivery latency for safety messages. The retransmission algorithm is evaluated in static scenarios using computer simulations and is shown that the algorithm can adapt to channel conditions through an exposed dynamically adjustable parameter. An alternative formulation of the algorithm containing dynamic transmission range control is also considered, but computer simulations show that it is not viable. Finally, suggestions on strategies to dynamically adjust this retransmission parameter are discussed.

4.2 Packet retransmission algorithm

The packet retransmission algorithm presented in this chapter aims to maximise the gain in coverage area while minimising the interference caused, thereby controlling the interference problem inherent in flooding based forwarding algorithms. Similar to the other flooding based algorithms, this algorithm exploits spatial diversity, enabling stations that were unable to receive the original packet to receive a forwarded copy from another station. Instead of using *a priori* knowledge of the network to minimise interference (which introduces overhead in establishing and maintaining this knowledge), this algorithm makes retransmission decisions based on a metric that takes into account the ratio between benefit and cost, *i.e.* additional coverage-to-interference ratio. Unlike Briesemeister *et al.*'s greedy algorithm [84], which uses station distance as the metric thus optimises hop count but generates interference outside the target area, the benefit-to-interference metric not only reduces interference outside the target area, but also naturally choose the furthest station if needed.

A few practical assumptions were made to enable implementation of the algorithm. First, the required coverage area of a packet is represented by a unit disc, with the centre and radius specified in the packet header. It is assume that the source station is close enough to the intended recipients. The header should also contain an expiry field to limit the extent of flooding by preventing the retransmission of stale packets. While the packet itself does not contain any history of whether or by whom the packet has been forwarded, each station may keep track of its received packets.

4.2.1 Retransmission metric

The retransmission metric used in the algorithm is based on the ratio of the area of additional coverage ("gain") to the area of interference, and is computed by considering the geometry of overlapping coverage areas. To simplify the metric, both unit-disc propagation, and uniformly distributed stations are assumed. This

Table 4.1: Symbols used in the retransmission metric model

R	Required range of packet
S	Range of retransmissions
r	Estimated range of packet
d	Distance between source and relay
A_D	Estimated area receiving duplicates packets
A_G	Estimated area receiving retransmission but not first transmission
X	An adjustable "weight" parameter

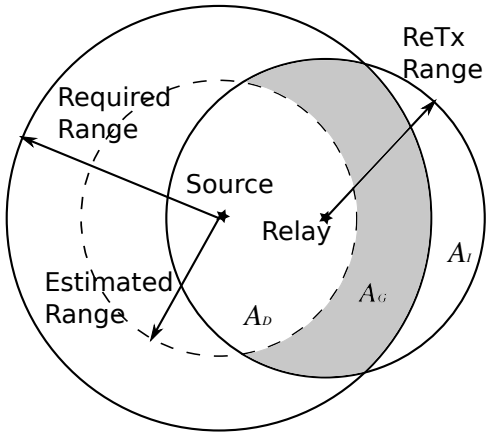


Figure 4.1: Derivation of retransmission metric — Required range is specified in the packet header; Estimated range is determined based on the packet's received signal strength; "ReTx range" is the transmission range of relay calculated based on its transmit power; A_D is the area receiving duplicate packets; A_G is the area additional coverage due to retransmission; and A_I is the area unnecessarily receiving the packet

simplification allows the number of stations receiving useful retransmissions and those being interfered with to be estimated assuming constant station density. Because the metric uses the ratio between these two counts, the station density cancels out in the function and becomes irrelevant. Table 4.1 lists the symbols used in this model. Figure 4.1 is a graphical representation of this geometry.

In this model, the following constraint must hold true for all real systems:

$$r \geq d > 0$$

The transmitter is assumed to use half-duplex radio, and does not overhear its own transmissions. $d > r$ represents the case where the relay is located outside the range of the original transmission, and would not have heard the packet. If

$d = 0$, the transmission could not have been received by any station.

The definition of A_D can be simplified by constraining the system to:

$$R \geq r \geq d > 0$$

For cases where $r > R$, the original transmission will reach further than required, thus $A_G = 0$. In this scenario, no retransmissions can improve the reception of the packet.

Using the geometry of intersecting circles [178]:

$$A_D = \begin{cases} r^2 \cos^{-1} \frac{d^2-S^2+r^2}{2dr} + S^2 \cos^{-1} \frac{d^2+S^2-r^2}{2dS} & \\ \qquad -\frac{d}{2}\left(4S^2 - \left(\frac{d^2+S^2-r^2}{d}\right)^2\right)^{\frac{1}{2}} & d \geq |S-r| \\ \pi r^2 & d < S-r \wedge S \geq r \\ \pi S^2 & d < r-S \wedge S < r \end{cases} \tag{4.1}$$

$$A_G = \begin{cases} R^2 \cos^{-1} \frac{d^2-S^2+R^2}{2dR} + S^2 \cos^{-1} \frac{d^2+S^2-R^2}{2dS} & \\ \qquad -\frac{d}{2}\left(4R^2 - \left(\frac{d^2-S^2+R^2}{d}\right)^2\right)^{\frac{1}{2}} - A_D & d \geq |S-R| \\ \pi R^2 - A_D & d < S-R \wedge S \geq R \\ \pi S^2 - A_D & d < R-S \wedge S < R \end{cases} \tag{4.2}$$

Finally, the metric for the retransmission decision can be easily calculated from these areas as the ratio of the area benefited to the areas receiving interference:

$$M = \frac{X A_G}{\pi S^2 - A_G} \qquad X \in \mathbb{R}^* \tag{4.3}$$

Here, X is the externally-adjustable configurable parameter related to the tolerance of interference. X can be any non-negative real number, with higher X equating to a higher tolerance to interference (stations are allowed to interfere more). Retransmissions can be turned off by setting X to 0, and if X is very large,

all stations will always retransmit. Using this metric, stations with higher M is better positioned to retransmit the packet because they have a better coverage-to-interference ratio. It is intended that retransmissions be permitted only for $M \geq 1$ (benefit is at least some threshold multiple of interference).

In the implementation of this algorithm used in chapter, stations estimate the range of each received packet (new or retransmitted) using the received signal strength. By assuming that the packet was sent from the centre of the target area and all attenuation was due to path loss, the areas above can be calculated, with the total area of interference being the difference between the coverage area of the retransmission and A_G (the areas receiving retransmitted packets but not the original).

4.2.2 Retransmission algorithm

The retransmission metric relaxes the need for neighbourhood knowledge when making the retransmission decision, but it still needs global coordination to correctly prioritise potential relays. To relax the need for explicit coordination overhead, a shared knowledge — time — is used. By assuming the difference in the speed of the real-time clock is insignificant (*i.e.* all vehicles agrees on the length of the same time period; time difference is irrelevant), and all stations can sense the transmission of another, one can use a delay-based algorithm to coordinate the transmission order.

$$delay = \frac{T_{pkt}}{M} \tag{4.4}$$

In this retransmission algorithm, a delay-based scheme similar to the prioritising scheme in M_GeRaF [110] and in Briesemeister *et al.*'s work [84] is used. Once a packet is received and forwarded up to higher layer, a retransmission delay is calculated based on the retransmission metric (M) using Equation 4.4. This function maps M to a delay value such that the delay for highly desirable stations (high M) tends to 0. T_{pkt} is the maximum lifetime of a packet. When

Algorithm 4.1 Retransmission algorithm

```
function RECEIVE_PACKET(pkt)
    if pkt has never been seen then
        Pass pkt to higher layer
    end if
    if a version of pkt (pkt') is in Tx queue then
        Estimate range of pkt based on received signal strength
        if pkt can propagate further than pkt' then
            Drop pkt' from the Tx queue
        else
            Ignore pkt
            return
        end if
    end if
    if pkt has never been seen or pkt' was dropped then
        M = CALCULATE_METRIC()
        t_delay = CALCULATE_DELAY(M)
        if pkt is not yet expired after t_delay then
            Add pkt to transmission queue with delay = t_delay
        end if
    end if
end function
```

$M < 1$ (when the area ratio is below the minimum specified), the delay exceeds the packet lifetime thus disabling retransmissions. A packet will be retransmitted after the calculated delay if a copy of the same packet with a higher received signal strength (hence can propagate further) was not received during the delay period. To avoid unnecessary transmissions, packets are not added to the retransmit queue if the calculated delay causes the retransmission to be scheduled after the packet has expired. If a copy of the packet with a higher received signal strength was received during the delay period, the packet is dropped from the transmit queue, and the new copy of the packet will go through the retransmission algorithm to determine a new delay. Algorithm 4.1 outlines the retransmission algorithm.

For the simplicity of implementation in the evaluation, the retransmission algorithm is independent of the MAC and link-layer mechanisms. When the algorithm "retransmits" the packet, it sends the packet down to the link layer. The packet is then subjected to the normal MAC delays (backoff, IFS, *etc.*). This

algorithm does not use the IEEE 802.11e mechanism for prioritising packets.

However, implementing this algorithm above the MAC layer without cross-layer notification of MAC state changes can cause clustering of delayed packets — many packets with different delay values may be submitted to the MAC transmit queue during the reception and transmission of another packet, and the algorithm is unable to remove packets currently in the MAC transmit queue. This clustering can overflow the MAC queue and can greatly increase the contention on the channel. A production-grade implementation would benefit from being advised of state changes at the MAC layer (thus able to pause the delay timer when channel is busy), and should be able to drop packets from the MAC transmit queue.

4.2.3 Range adaptation

In addition to the algorithm described above, this chapter also investigates a variant of the algorithm that moderates the range of the retransmissions. It can be hypothesised that interference may be further reduced by moderating the retransmission power, thus reducing the area where duplicate packets are received (A_D), and may also reduce the area where irrelevant packets are received (area outside the intended area) in certain cases. However, a reduction in range may also decrease the area of additional coverage.

Assuming each station can only transmit at discrete power levels, a simple retransmission range adaptation scheme is considered where each station, instead of calculating the retransmission metric for a single retransmission range, calculates M for each available transmit power. The transmission power (hence the range) that yields the highest M is used for the retransmission, with retransmission delay calculated based on the highest M.

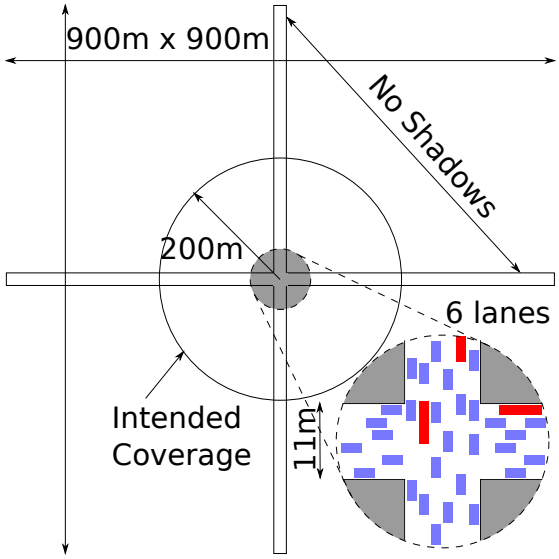


Figure 4.2: Vehicle layout of intersection scenario — Tagged sender at the centre of circle

4.3 Performance evaluation

This algorithm is evaluated using the network simulator ns-3.9. In these simulations, vehicles are arranged in two intersecting, 900 m long, 6-lane road segments (3 lanes each way for each road segment). Since the required coverage area is a 200 m radius circle at the centre of the field, and the maximum transmission range of each vehicle is at most 200 m, this field is sufficiently large to capture all expected interference and to avoid edge effects. Given station mobility is negligible for the duration of the short VANET safety messages, the network can be approximated by as a static topology to avoid the ripple effects caused by stations entering and leaving the coverage area during a simulation.

Vehicles in the simulation are distributed linearly along each lane, with the spacing between consecutive vehicles following a Poisson distribution. This produces lanes of vehicles with approximately the specified vehicle density. In the cases where two vehicles overlap each other, their positions are adjusted to the minimum vehicle separation. Two types of vehicles are modelled: cars (5.5 m×2.5 m, no shadows) and heavy vehicles (12 m×2.5 m, attenuates all packets that travels into/out of/across it by 20 dB). Transmitters in these scenarios generate CCA beacons at 10 pkt/s. For each vehicle density setting, 10

Table 4.2: Layout parameters

Vehicle densities	{10, 50, 75, 100} veh/km/lane
Proportion of HV	10%
Lanes per road	6
Road length	900 m
Intended coverage radius	200 m
Position of source	Centre of intersection
	80 m south of intersection
	Centre of straight road

situations are tested with results presented being the average of these situations. Figure 4.2 shows a typical vehicle layout for the centre-of-intersection scenario. Specific parameters used to generate the vehicle layouts are listed in Table 4.2.

Three different sets of vehicle layouts are simulated — the tagged sender at the centre of intersection, sender 80 m south of the intersection and straight road (no intersections). The centre-of-intersection layout represents the case when the highest channel contention is near the source; sender offset from intersection layout represents the case where station distribution is not homogeneous and many stations need relays to receive the packet; and the straight road layout has homogeneous station distribution, and building shadow is irrelevant. A total of ten layouts were generated for each combination of the layout parameters.

A log-distance path loss model combined with the Nakagami fast fading model is used, in addition to the vehicle shadowing described above as the radio propagation model. A building shadow of -30 dB is also simulated if the line-of-sight between the two communicating vehicles crosses a building, assuming the building sits on the edge of the road segments. Table 4.3 details the channel characteristics simulated. Channel parameters used are either the default values (Rx and CS Thresholds, Receiver Noise), values used in many existing literature (Log-Distant Exponent, Nakagami parameter m, Antenna gain) [125, 179] , or fundamental computed values (Log-Distance reference loss using the Friss equation [180]). Timing and rate parameters are as per IEEE 802.11 specifications. Radio shadow depths are assumptions based on measurements in [55–61]. A

Table 4.3: PHY, channel and MAC characteristics

Antenna Gain (Tx and Rx)	2.512 dB
Rx Threshold	-95 dBm
CS Threshold	-99 dBm
Log-Distant Exponent (γ)	2.0
Log-Distant Ref Loss (at 1 m)	-47.8588 dB (Friis loss, 5.9 GHz)
Nakagami parameter (m)	5.0
Receiver Noise	0 dB
Attenuation across HV	-20 dB
Building Shadow	-30 dB
Max transmission range	{50, 100, 150, 200}m
Transmission rate	6 Mbps
Transmission bandwidth	10 MHz
SIFS	32 μs
Slot Time	13 μs

10 MHz OFDM channel is used for the PHY model, under an NQoS 802.11 MAC layer. The application modelled transmits raw MAC frames, with no management algorithms for any layer above MAC being active, except for the retransmission algorithm under test. The 802.11 PHY model used does not account for the packet capture effect.

Two groups of simulations are performed — one group with only the tagged station transmitting, used for determining the performance in near-optimal (uncongested) situations; the other with all stations transmitting regularly, approximating the intended use-case of Cooperative Collision Avoidance systems (CCA). Parameters used for these scenarios and algorithms are listed in Table 4.4. The packet size used is based on a IEEE 802.11 data frame [44] + 20 bytes of beacon data [33]. The packet rate used is the typical value proposed for CCA applications [19, 22, 23, 134].

In the CCA use-case, even though all stations transmit packets in the simulation, only the packets generated by the tagged station are tracked in the simulation. For each layout, background beacons are generated at 10 pkt/s (as proposed for many VANET applications) with start times randomly drawn from a uniform distribution over $[0, 100)$ ms. A sample of 20 tagged packets is taken

for each algorithm under test. The performance measures taken include the number of relevant packets successfully received before expiry, the number of packets that arrived expired, the number of redundant or irrelevant packet received, and the amount of time the radio interface is in their various states (channel busy, receiving and transmitting).

The behaviour of the metric-based algorithms are compared to some well-known approaches that have different strengths and weaknesses. AFR-CS [22] is the initial strategy proposed when the issue of reception ratio was first highlighted and represents a baseline mechanism. In AFR-CS, the sender resends each packet k times in order to combat fast fading. Furthest successful station [84] implements Briesemeister *et al.*'s greedy algorithm commonly used for geographical routing. A scheme where only heavy vehicles retransmit is also evaluated because shadows in our scenario are cast by heavy vehicles. This study does not consider the effect of reduced attenuation amongst heavy vehicles observed by Boban *et al.* [89], which was published after this study was completed. Comparing our algorithms (with and without range adaptation) with these casts wide spectrum light on the performance of our approaches.

4.4 Results

4.4.1 Connectivity and interference

Connectivity of the network due to the operation of the retransmission algorithms can be observed be looking at the packet reception ratio (PRR) for single-source scenarios. Here, PRR is the proportion of tagged packets that were correctly received before packet expiry, aggregated over the stations within the required coverage area only. A packet is considered correctly received if at least one copy of the packet (original or retransmitted) is received.

The amount of interference introduced by the algorithm can be observed by looking at the channel busy time (CBT). CBT is the total amount of time a

Table 4.4: Algorithm and scenario parameters

Packet Rate	10 Hz
Packet lifetime	100 ms
Packet size	54 octets (incl. all headers)
Adaptive range increments	10 m/level (up to max Tx range)
Retransmission Metric param X	{0.5, 1, 2, 3, ..., 10}
Number of transmitters	{Single station, All stations}

station's radio interface is either transmitting, receiving or has otherwise sensed a carrier on the channel. The CBT is also only aggregated over the stations within the required coverage area to minimise edge effects. In the figures, the CBT is displayed as a percentage of the total lifetime of the tagged packets (100 ms $\times$ 20 packets = 2000 ms). It should be noted that the CBT by itself may not be representative of the channel contention introduced by these algorithms — consider the case where hundreds of relays wanting to access the channel within a very short fraction of the message period, causing very high contention but the CBT value will be low. This case results in a low PRR due to contention, even though the CBT might also be low. In terms of "interference" as defined in this chapter, this is actually a low-interference scenario as the multiple concurrent receptions of packet (collisions) will only cause the radio interface to be blocked once.

Figures 4.3a and 4.3b show the overall PRR and CBT aggregated over the ten centre-of-intersection scenarios for each vehicle density tested, with the ratio of maximum transmission range to required radius (S) equals to one. These scenarios represent the cases where, in the absence of fast fading, shadowing, collision and without retransmissions, all vehicles within the coverage area should receive all tagged packets and no vehicle outside the area should receive any packets (*i.e.* unit disc propagation). Figures 4.3e and 4.3f show the performance when S is 0.5, which is a multi-hop broadcast scenario. Figures 4.3g and 4.3h represent the scenario where a packet needs to be forwarded at least 3 times before arriving at the edge of the required coverage area.

As expected, increases in vehicle density cause the PRR to decrease in the

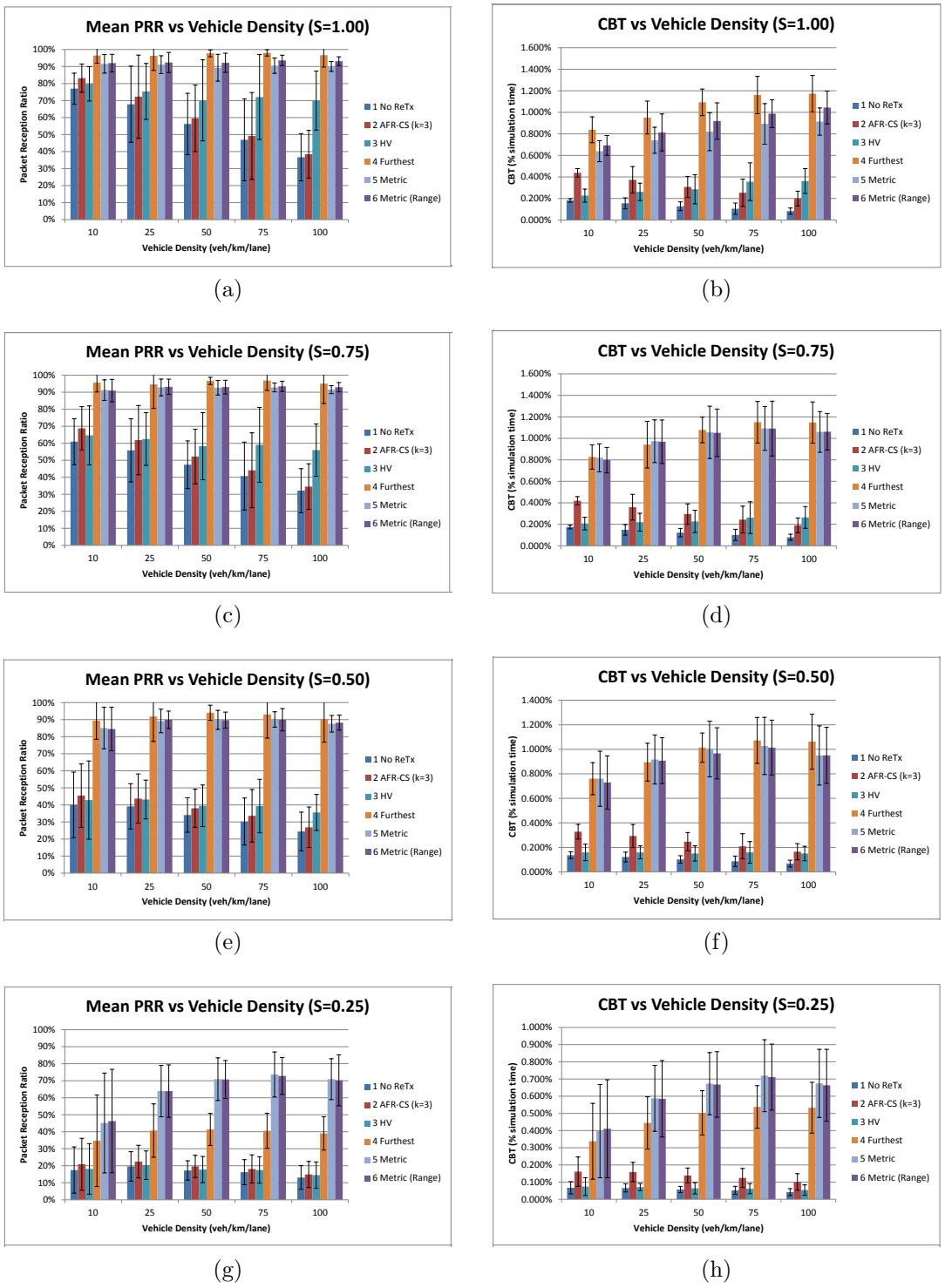


Figure 4.3: Results of centre-of-intersection scenario, single transmitter, error bars represents two standard deviations, S is the ratio of the maximum transmission range to required coverage radius. The graphs plot the means and 2 standard deviations of the mean value in each test. Parameter X for the metric cases are chosen to give the best PRR and is common only between corresponding columns in the PRR and CBT graphs. PRR graphs on the left, and the corresponding CBT on the right for $S \in \{1.00, 0.75, 0.50, 0.25\}$.

no-retransmission case (Figure 4.3, left plots), due to the increased number of heavy vehicles between the sender and receiver. PRR is slightly improved when using AFR-CS [22], which overcomes fast fading but not the shadowing caused by the heavy vehicles. As the vehicle density increases, the improvements due to AFR-CS diminishes as more packets are lost due to shadowing than fast fading.

It is observed that, in high enough vehicle densities, the heavy vehicle-only scheme ("HV") performs rather well. This scheme's success can be attributed to the low proportion of heavy vehicles in the test scenarios, thereby heavily restricting the set of relays. At low local vehicle densities (*i.e.* the number of vehicles within one transmission radius), the performance is poor, probably due to the lack of potential relays within range. This lack of relays can be seen in the CBT graphs, which shows that at low local densities, the scheme barely uses the radio channel.

The furthest successful station algorithm also tends to produce an approximately 5% improvement in PRR over the two benefit-interference based algorithms in single source scenarios. However, when the transmission range is very short ($S = 0.25$, Figure 4.3g) its performance is observed to be poor even though the algorithm is designed to minimise hop count. Its poor performance can be attributed to the way delays were calculated in this algorithm. In Briesemeister *et al.*'s algorithm [84], delays were calculated linear to the distance of the relay from the source, with no considerations paid to how far the packets can actually propagate. This preference of further stations introduces a very large retransmission delays when the transmission range is very short compared to the required range (in this case, a quarter of the required coverage radius), causing many packets to expire before retransmission. This effect can be confirmed by observing the corresponding low CBT values in Figure 4.3h, which suggest that packets aren't actually being forwarded.

Both benefit-to-interference metric based algorithms performed similarly, achieving a good (but not perfect) reception ratio. It is noteworthy that the

amount of interference caused is comparable (and often lower than) the furthest successful station algorithm. It is acknowledged that the metric based retransmission algorithm is never designed to achieve 100% PRR — given any finite value for parameter X, it is always theoretically possible that all potential relays will not have a high enough benefit-to-interference ratio to allow retransmissions, even if there is a significant area that is estimated to not be able to receive the packet.

Additionally, one can see that the value of the chosen parameter X for all of the single-transmitter scenarios are high. High X encourage retransmissions, thus increases the probability that the message will be retransmitted quickly. It also has the effect of increasing the amount of interference introduced. Figure 4.4 shows the effect of varying this parameter.

Looking at Figure 4.4, it is obvious that increasing the value of parameter X increases both the CBT and PRR for any given scenario. It is also important to observe that after a certain threshold for X, the effect of increasing X diminishes, but at the same time, the reduction in CBT growth is much slower than the reduction in the PRR growth. This suggests that covering the last few percent of stations gets increasingly harder and may not actually be viable. The next section investigates the effect of the increasing CBT on PRR (due to the application of retransmission algorithms) in the CCA application context.

4.4.2 Algorithm performance in an application context

The performance of these algorithms were also investigated within the Cooperative Collision Avoidance context, assuming multi-hop beaconing is permitted. The following three performance metrics — packet reception ratio (PRR), channel busy time (CBT) and packet delay are evaluated. Packet delay is defined as the time between a packet's generation and its first successful reception at that station (receptions of duplicates are ignored). Delay values are collected at each station and is inclusive of the time needed for transmission (220 μs), all retransmission

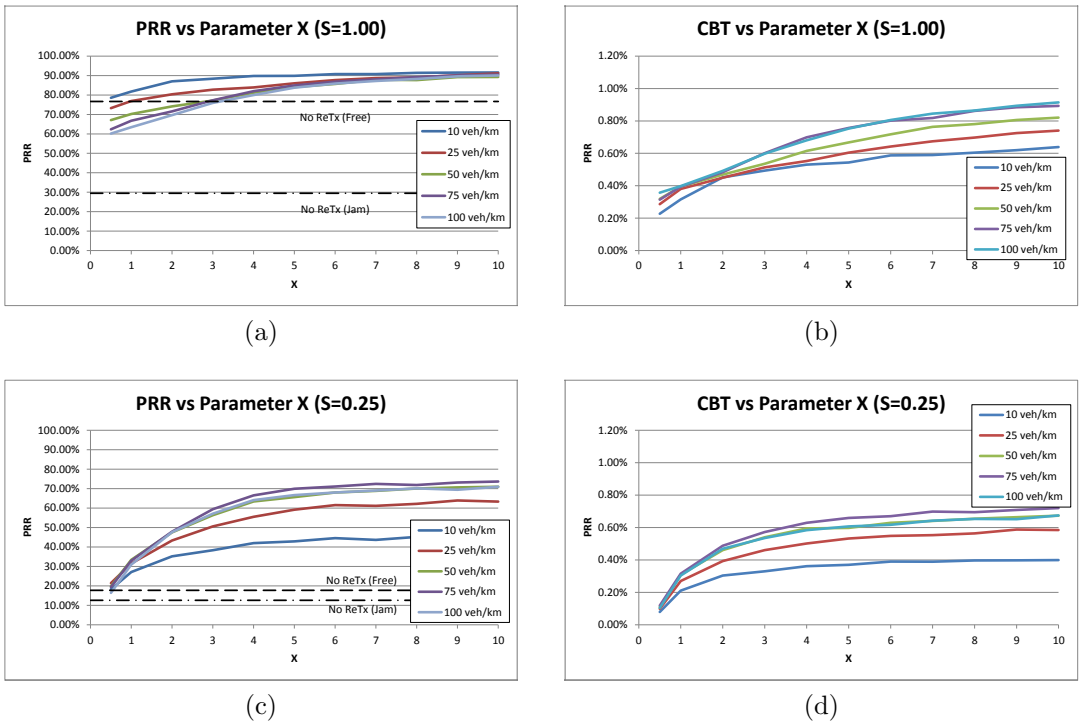


Figure 4.4: Effect of varying parameter X in centre-of-intersection scenario, single transmitter. S is the ratio of the maximum transmission range to required coverage radius. PRR for no retransmission at high and low vehicle density marked for comparison. PRR graphs on the left column, and CBT on the right for $S \in \{1.00, 0.25\}$. For clarity, only the mean values are plotted. Two-standard deviations (95% CI) values for packet reception are typically higher in sparse layouts than dense layouts and decreases with increasing X. Their ranges are: (a) 5–11% at 10 veh/km/lane and 3–14% at 100 veh/km/lane; (c) 13–32% and 6–15% respectively. CBT confidence intervals are approximately the same, all decreases with increasing X. In (b) two-standard deviations range 0.05–0.1% for sparse layouts, and 0.06–0.1% for dense layouts; (d) ranges are 0.05–0.3% and 0.06–0.2% respectively.

delays, and all MAC queuing delays. The simulations terminate after 2000 ms, when all packets not yet received will be expired. Figure 4.5 shows the overall PRR and CBT across the different scenarios using the centre-of-intersection layouts.

Similar to Section 4.4.1 and as expected, increases in vehicle density causes the PRR to decrease in the no-retransmission case (Figures 4.5a and 4.5e), and slight improvements in PRR was observed using AFR-CS [22]. Both of these algorithms produce little interference, thus performance degradation due to station density were not observed in the tests conducted. The performance of the heavy vehicle-

Figure 4.5: Results of centre-of-intersection scenario, all stations transmitting. Graphs plot the means and 2 standard deviations of the mean values in each test. S is the ratio of the maximum transmission range to required coverage radius, saturation marks the theoretical maximum utilisation. Parameter X for the metric cases are chosen to give the best PRR and is common only between corresponding columns in the PRR and CBT graphs. PRR graphs are on the left, and the corresponding CBT on the right for $S \in \{1.00, 0.75, 0.50, 0.25\}$.

only scheme ("HV") is also unchanged compared to the single transmitter case for the same reason.

The furthest successful station algorithm ("Furthest") [84] is only effective at low vehicle densities and degrades rapidly as vehicle density is increased. At 10 vehicles/km/lane, the channel contention is still very low and the algorithm provides sufficient redundancy by triggering at least one retransmission for every packet and their retransmissions until packet expiry. As station densities increase (even slightly from 10–50 vehicles/km/lane), it can be seen that the furthest-station algorithm cause packets to not be received at all. This indicates that this algorithm is actually detrimental to packet reception as the density scales up. By observing the channel busy time (Figures 4.5b and 4.5f), one would conclude that these algorithms cause too much interference, either by triggering too many retransmissions or triggering them at suboptimal locations.

Figure 4.5b shows the amount of interference caused by the various retransmission schemes as measured by the channel busy time. As the vehicle densities increase, it can be seen that the CBT for the no-retransmission case also increases due to the increased load (each station transmits 10 packets per second). It is noted that the increase is not linear. This suggests that higher densities also increases the probability that two stations would pick the same backoff slot to transmit, and therefore increase packet collision events.

The furthest successful station algorithms causes an approximately 7-fold increase in CBT compared to the no retransmission case, even in the lowest density scenario (*i.e.* each packet is retransmitted at least 7 times). This level of retransmission is unsustainable, and results in an unacceptable amount of packet collisions and lengthening of the MAC queue delays, causing packets to expire while still in the transmit queue at higher densities. The "Saturation" line marks the theoretical maximum channel utilisation for frame duration = 220 μs and DIFS = 58 μs — the CBT for this algorithms approached and even exceeded the maximum channel utilisation at densities higher than 50 vehicles/km/lane.

Utilisation in excess of the theoretical maximum can occur due to the hidden terminal problem.

Figures 4.5a and 4.5b compare the metric based algorithms with the others. It is important to note that the graphs show the results of the metric based algorithms, with parameter X chosen to give the best PRR at that vehicle density. In terms of PRR, the results showed that the metric based algorithms (with and without range adaptation) provide some improvements/ over AFR-CS at lower vehicle densities. At high densities, these algorithms still performed poorer than AFR-CS, but the degradation is slower than the other algorithms due to its lower interference.

It is also noteworthy that the minimum value of X that had been tested is 0.5. The performance of the algorithm with X set to 0 is identical to the performance of the case without any retransmission ($X = 0$ disables retransmissions). Therefore, these retransmission algorithms should perform no worse than the no-retransmission case as long as X is adjusted appropriately.

Figures 4.5e and 4.5f show the benefit of the algorithm in terms of spatial diversity. These results correspond to the case where the source can only reach half its required radius, hence AFR-CS is ineffective in providing any improvements. Here, the furthest successful station algorithm provides the best PRR in low vehicle densities because this is the situation it was designed for. The metric based algorithm is also able to produce the multi-hop behaviour in these situations, as evident by the high PRR at low vehicle density. In addition, the lower interference it introduces enabled it to degrade slower than the other schemes (it even provided better PRR over AFR-CS at 100 vehicles/km/lane) whilst the furthest successful station algorithm quickly saturated the channel, causing packets to not be received at all.

In terms of range adaptation, the results in Figure 4.5 do not show any significant differences between the two algorithms. The range adapted algorithm appears to perform slightly worse than the non-range adapted version, with slightly

decreased PRR and slightly increased CBT. It is conceivable that the range adaptation as implemented may be able to reduce the size of the area experiencing interference from the relay, but at the same time increases the channel contention as the improvement in the metric causes the delays to decrease across the network. Given the minute differences observed and the increased implementation complexity, range adaptation in its current form is not viable.

4.4.3 Effect of vehicle layouts

Three types of vehicle layouts were simulated — tagged sender at the centre of the intersection, tagged sender 80 m south of the intersection (the "offset scenario") and the tagged sender along a straight road. Figures 4.6 and 4.7 compare the performances of the algorithms at each layout for $S = 1.00$ and $S = 0.25$ respectively.

Overall, the results for each layout with the same transmission range are similar, showing quite similar trends, with the straight road achieving slightly better performance across all measures for all schemes. This is possibly due to the lower vehicle density and the transmissions not being affected by building shadows.

When the maximum transmission range is high, the metric based algorithms continues to provide effective interference minimisation, achieving better PRR than most other retransmission schemes. The heavy-vehicle only scheme performed well at high densities due to its much smaller set of relays, but it is unable provide much improvement at low densities, regardless of the vehicle layout.

The offset scenario illustrates the problem with the range adapted retransmission algorithm, where the higher interference is more pronounced. The vehicle layout in this scenario is highly non-uniform — there are more vehicles closer to the intersection (which is not near the source). Therefore the number of potential relays are higher near the intersection. The range adaptation allows all these potential relays to adjust its output power, boosting the potential relay's

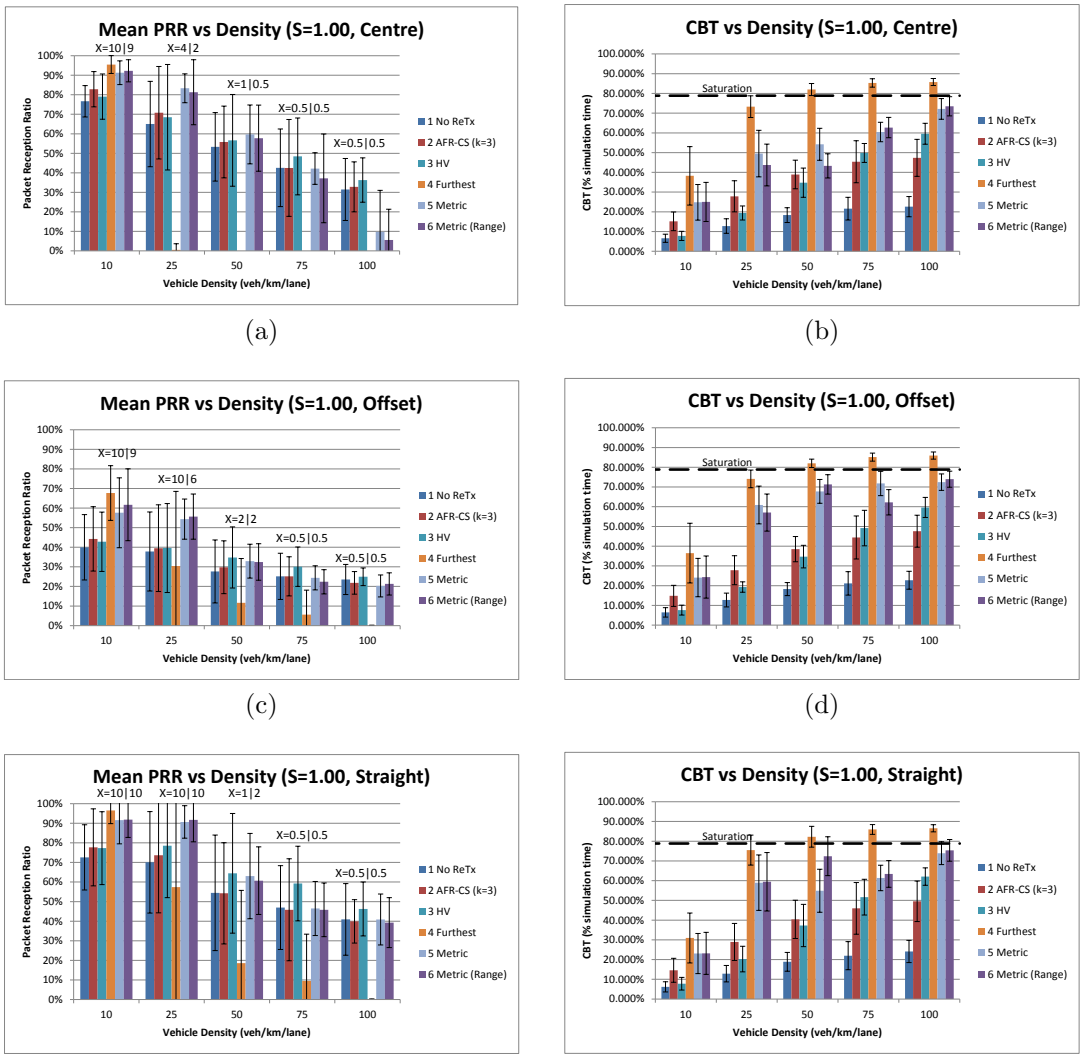


Figure 4.6: Results with all stations transmitting, error bars represent two standard deviations, all target stations should receive packet if under unit-disc propagation. Parameter X for the metric cases are chosen to give the best PRR and is common only between corresponding columns in the PRR and CBT graphs. First row is centre-of-intersection case, second row is 80 m south of intersection, and last row the straight road scenario.

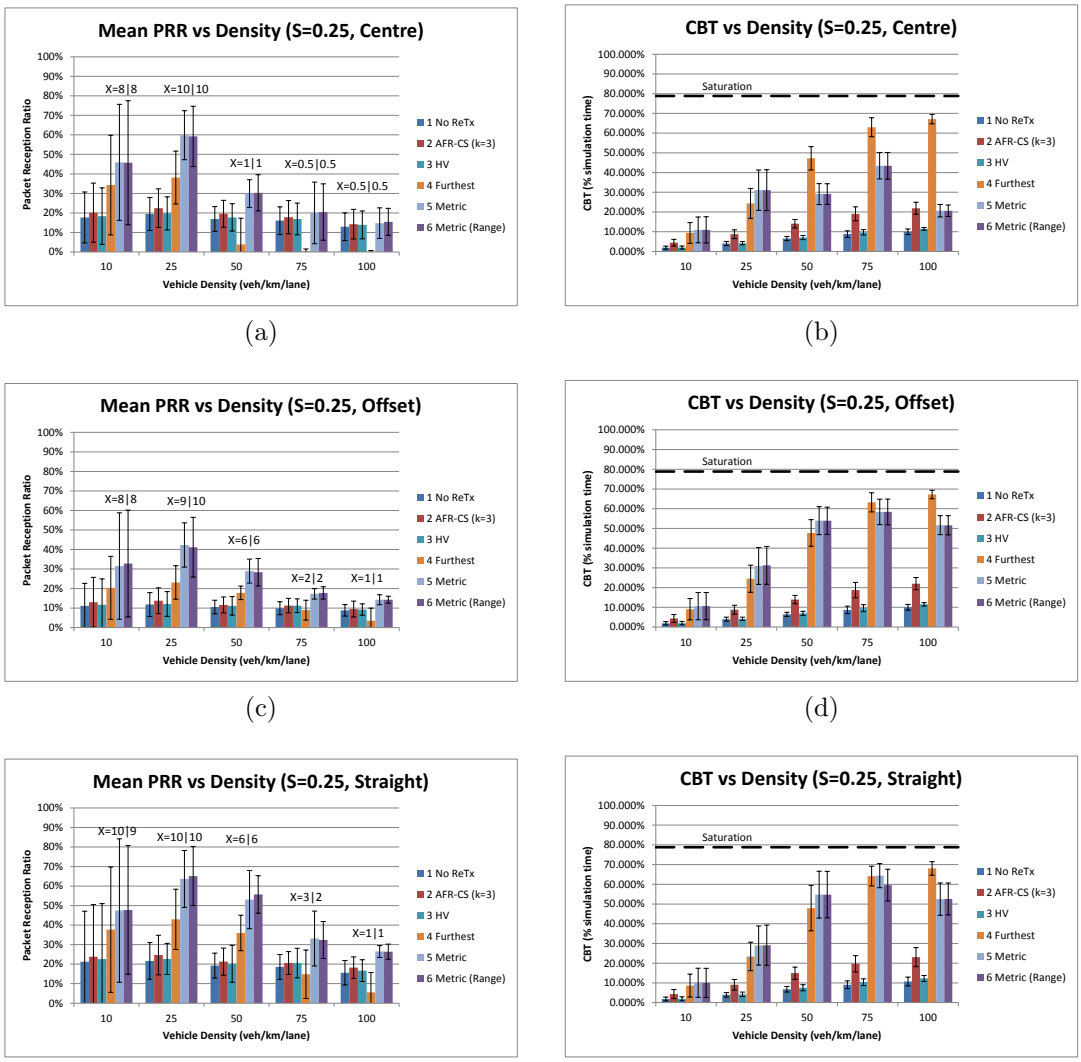
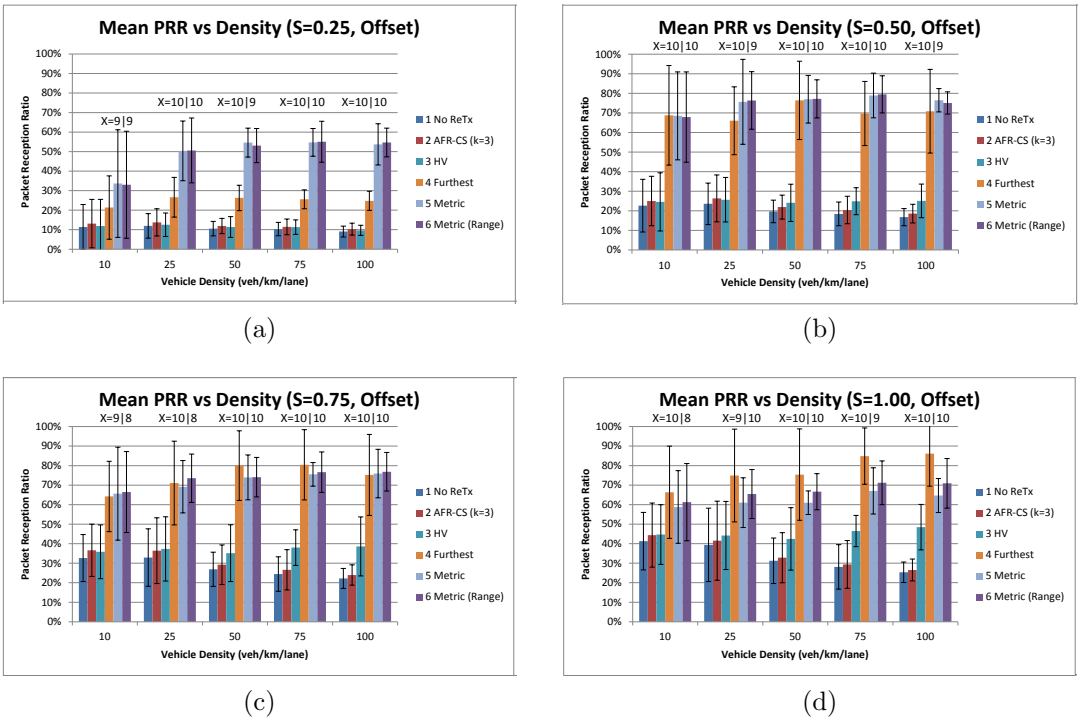


Figure 4.7: Results with all stations transmitting, error bars represents two standard deviations, maximum transmission radius is one quarter of the required distance. Parameter X for the metric cases are chosen to give the best PRR and is common only between corresponding columns in the PRR and CBT graphs. First row is centre-of-intersection case, second row is 80 m south of intersection, and last row the straight road scenario.



(a) (b)

(c) (d)

Figure 4.8: Results with a single transmitter, error bars represents two standard deviations. Parameter X for the metric cases are chosen to give the best PRR. Graphs plots PRR for maximum transmission range ratio $S \in \{0.25, 0.50, 0.75, 1.00\}$.

desirability to retransmit. Figure 4.7d shows that more interference is introduced (probably due to more stations deciding to retransmit), but the PRR is not much better (if at all) than the non-range adapted version (Figure 4.7c).

The non-uniform distribution of vehicles also reduces the retransmission algorithms' ability to provide good coverage. Figure 4.8 shows the PRR of the various algorithms with only a single transmitter — channel congestion is not an issue in these tests. Figures 4.8a and 4.8b show that the algorithm performs fairly well in scenarios that obviously require forwarding by the furthest station. When the maximum transmission range increases, the potential benefits (in terms of area) for the relays diminish causing a reluctance in retransmission. However, since there are many vehicles near the edges of the coverage area, non-retransmission causes a higher drop in PRR than scenarios with uniform distribution. Figures 4.8c and 4.8d suggest that the metric based algorithms had reached its peak PRR (parameter X is at the highest value tested). This problem of estimating benefit

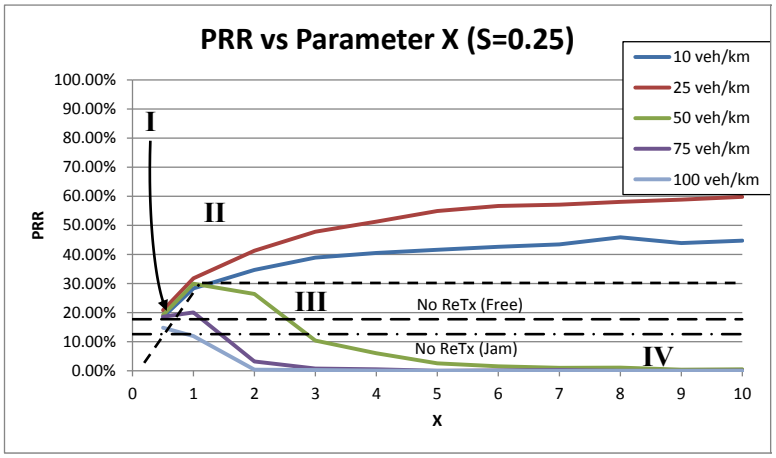
by area instead of station count is similar to the analogy of having telecommunication companies providing 98% mobile coverage in a country in terms of area, but the 2% that are not covered is located in the major cities because they did not consider population distribution. This problem cannot be resolved unless the potential relays can also estimate the distribution of vehicles as well.

4.4.4 Variation of parameter X in an application context

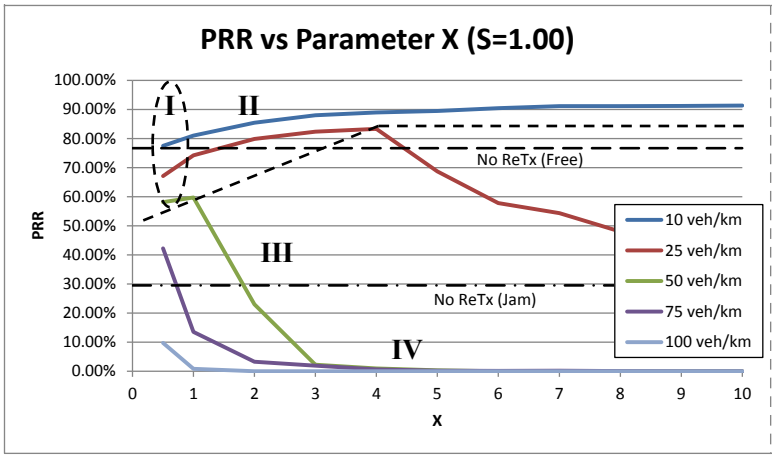
The results above showed the performance of the algorithms with the algorithm parameter X chosen to provide the best PRR for a specific vehicle density. Figure 4.9 shows the effect of varying parameter X, and shows the regions of interest. Not all of the regions were observed when S is higher due to the granularity of the variations in X.

There are four regions in general, numbered I, II, III and IV. Region I represents the range of X where no retransmission was triggered by the algorithm (hence the performance of the algorithm is identical to the no retransmission case). With reference to Figure 4.9a, only the end of this region is observed, and only in the "free" (10 veh/km/lane) case at $X = 0.5$. Region II represents the values of X where improvements can be gained by encouraging more retransmission (*i.e.* increasing X). This correlates to the range of X with positive slope in the PRR curve. Region III is the region where increasing the number of transmissions causes performance to deteriorate due to collision and excessive delays (*i.e.*the region with negative slope). Finally, Region IV represents the region where no successful reception can be expected, either due to collision or delays.

From Figure 4.9a, it is observed that the optimal value of X (the value of X between Regions II and III) decreases non-linearly as vehicle density increases. When comparing between Figures 4.9a and 4.9b, it is also observed that as S increases, optimal X also decreases.



(a)



(b)

Figure 4.9: Performance of the metric (without range adaptation) algorithm vs parameter X, in the centre-of-intersection scenario, all stations transmitting. Ratio of transmission range to required coverage radius (S) ranges between 0.25 and 1. Error bars omitted for clarity. The PRR without any retransmissions (at densities of 10 and 100 veh/km/lane) are shown for comparison. (a) packet reception ratio ($S = 0.25$); (b) packet reception ratio ($S = 1.00$);

4.5 Discussions

4.5.1 Behaviours and limitations of the algorithm

Using the metric presented, the retransmission algorithm optimises the proportion of additional coverage at each retransmission. The parameter X was introduced to control the level of tolerable interference, providing a lever to allow dynamic adjustment based on channel conditions. Under certain conditions, the metric based non-range adapted algorithm behaves similar to common algorithms:

- As X approaches 0, retransmission is discouraged, and the algorithm will not retransmit when X is 0;
- As X increases, retransmissions are encouraged while the delays between stations of different priority are reduced. As X approaches infinity, the algorithm becomes "Always retransmit immediately" (*i.e.* flooding);
- As maximum transmission range reduces past half of the required radius, the optimal location moves towards the edge of the transmission range (no risk of interfering outside the required area). The algorithm then behaves similar to the furthest successful station algorithm;
- As the maximum transmission range approaches the required radius, the optimal location moves towards the centre of the coverage area. For very high station densities, the algorithm approximates the AFR-CS algorithm (except the retransmission is made by a station very close to the source instead of the source itself).

The algorithm's flexibility allows it to adapt its behaviour as needed.

On the other hand, the metric assumes log-distance path loss as the only cause of signal attenuation. This simplified the calculation of coverage-vs-interference metric, but in reality, the actual additional station coverage vs interference can be very different to the calculated value. Consider the case where an obstruction

completely blocked transmissions in certain directions from a station but leave some paths unattenuated (*e.g.* at an intersection with a building that casts very deep shadow), the necessary relay stations on the same road segment will not be triggered to retransmit as they cannot sense any unexpected attenuation.

In addition, the algorithm uses a maximum interference tolerance (X) to limit the number of retransmissions. This has the side effect that the algorithm cannot guarantee 100% packet reception even in perfect conditions because the ratio of coverage-vs-interference will approach 0 as the original packet estimation approaches the required radius ("law of diminishing returns"). This means that for any chosen X, there will always be some "critical" estimated range, above which retransmission will not occur even if all stations know the packet will not be received by all intended recipients.

4.5.2 Determining parameter X

To provide the scalability demonstrated, the simulation results showed that the parameter X needs to be varied as vehicle density changes. The correct choice of the value of X is essential for the correct function of the algorithm. A low X discourages retransmissions except when it is highly advantageous (preferable for high vehicle densities), and will not retransmit at all when X is 0. A high X allows more stations to retransmit, but also reduces the effectiveness of the priority scheme, increasing the potential for packet collision (may be required for very low vehicle densities). The additional retransmissions can also increase packet delays if the retransmit queues are too large [181].

Based on the observations in Section 4.4.4, X should be a decreasing function of the number of one-hop neighbours. As the number of one-hop neighbours increases (by increasing transmission range or increasing vehicle density), the number of potential interferers increases. Since the number of actual interferers may be difficult to determine, one may be able to infer the level of interference/contention by observing the channel.

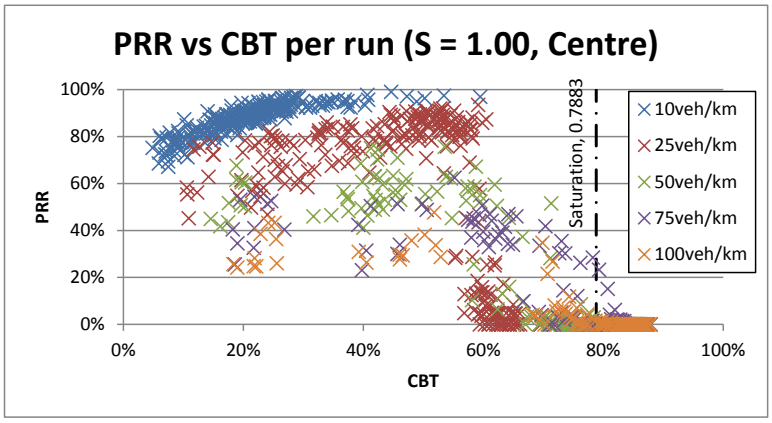


Figure 4.10: Measured PRR vs measured CBT, Centre-of-intersection scenario, all transmitting, $S = 1.00$. Graph plots the mean values of each test conducted.

Plotting PRR vs CBT (Figure 4.10) seems to indicate that there is a loose relationship between the two measurements. It appears that high CBT is associated with low PRR, but the mapping is not consistent. PRR by itself (or packet loss, as used in TCP) is also a poor indicator for channel contention in a wireless network since fading and shadowing also causes packet loss. Figure 4.10 suggests that in order to achieve an acceptable PRR ($\leq 50\%$), CBT should be kept below 60%. However, since the various CBT level were achieved using different algorithms (the figure plots data points from all algorithms and parameters tested), the CBT may not accurately reflect the actual contention experienced. One can see in Figure 4.10:

- PRR $> 70\%$ for CBT $< 40\%$;
- Non-reception at CBT above 70%
- PRR is scattered between 40% and 70%
- Transitions between these regions are quite distinct, possibly due to the granularity of vehicle density changes.

Another strategy is to moderate retransmissions based on an estimate of the local channel contention instead of simply using the CBT. Heusse *et al.* [141] presented a technique whereby the MAC layer adjusts its contention window such

that the observed idle slots is at the theoretically determined optimal. While this technique can improve packet reception and throughput for slightly delay-tolerant messages it cannot adjust the load offered to the MAC layer. A reactive strategy that moderates the retransmission parameter X may be able to achieve similar effect by moderating MAC-load instead of the contention window.

4.6 Conclusion

In this chapter, a cooperative geocast algorithm that determines whether to retransmit based on estimated coverage and interference was evaluated. The algorithm is shown to be effective in improving the dissemination of short packets over a prescribed area in the presence of both fast fading and shadowing. The algorithm uses a delay-based approach for prioritising packets, with the delay calculated based on a metric that accounts for additional coverage, redundant and irrelevant packet reception areas. When compared to selected alternative algorithms, the presented algorithm is effective in reducing the interference caused by retransmissions, and degrades slower as station density increases. Since the operation of the algorithm only requires local information and information already in packet header, minimal overhead is required.

In the next chapter, a novel method of estimating local channel contention is explored, paving the way for the dynamic adjustment of the retransmission algorithm parameter.

Overview

This chapter presents a passive channel contention estimation technique. As part of the development of this technique, a theoretical model of the broadcasting DCF is analysed to gain insight between channel contention and interframe idle period.

Contributions

- **I investigated the relationship between wireless channel contention and observed MAC-layer idle slot counts.** A Markov model of the MAC-broadcast DCF was constructed in this investigation. Numeric solutions to the model provides a mapping between the probability distribution of interframe idle slot counts and the channel contention in terms of the number of concurrent saturated stations. This mapping can be used by the MAC layer to estimate channel contention, in order to adjust MAC parameters and/or to provide feedback to upper layers for moderating the offered load onto the network.

- **I demonstrated and evaluated a passive technique for estimating channel contention using simple Bayesian inference.** Using the probability distribution computed from the Markov model, the technique of estimating contention through observing idle slots was compared to Bianchi *et al.*'s MAC-level contention measurement technique using computer simulations. I have shown that estimates from this technique converge to the scenario parameter quicker and is more accurate.

Publications

- **Tse, Quincy**, Si, Weisheng and Taheri, Javid, "Estimating Contention of IEEE 802.11 Broadcasts Based on Inter-Frame Idle Slots," in Proceedings of *Local Computer Networks Workshops (LCN Workshops), 2013 IEEE 9th Conference on*, pp. 120-127, Sydney, Australia:IEEE, 21-24 October 2013.

Chapter 5

Estimating contention of IEEE 802.11 DCF broadcasts without hidden terminals

5.1 Introduction

Chapter 4 introduces a retransmission scheme that can effectively balance the amount of interference generated by retransmissions, requiring no *a priori* knowledge of the network. However, the success of the scheme relies heavily on the correct dynamic adjustment of a parameter so that retransmissions are encouraged when the channel is free, but are suppressed when channel is busy. Even though the selection of the parameter may be done by observing the channel busy time (CBT), CBT itself is not a good measure of contention, and does not allow the station to quickly adapt.

In this chapter, an expression linking the number of idle slots between consecutive transmissions and the number of saturated stations is derived, based on a broadcast variant of Bianchi *et al.*'s Markov model [148]. Second, this relationship is exploited using Bayesian inference, observing the interframe idle slots in order to estimate the channel congestion level in terms of the number

Table 5.1: Notations used — all functions take parameters N and CW implicitly. These parameters are not marked on these symbols unless their values are unclear.

CW	Contention window size
N	Number of saturated stations (or equivalent)
P_k	Probability of choosing slot k from $[0, CW] \subset \mathbb{Z}$, distributed according to the solution of the Markov model
R_k	Probability of successfully receiving a transmission at slot k
R	Probability of successfully receiving a transmission
${}^{n}T_k$	Probability of n concurrent Tx at slot k
${}^{n}T_{k\|{}^{0}T_{k-1}}$	Probability of n concurrent Tx at slot k given no transmissions at slot $k-1$
T_k	Probability of the first Tx is at slot k
${}^{n}T$	Probability of n concurrent Tx
U_k	Probability of choosing slot k from $[0, CW] \subset \mathbb{Z}$, distributed uniformly
X_k	Probability of a collision at slot k
X	Probability of collision k

of saturated stations on the network. Here, a station is "saturated" if it always has at least a frame in its transmit buffer. Note that only the legacy Non-QoS IEEE 802.11 Distributed Coordination Function (DCF) is considered in this chapter. This technique can also predict the collision probability assuming ideal channels with no hidden terminals. Using computer simulations of the IEEE 802.11 DCF, estimators configured using this model are shown to be more accurate in estimating the channel contention, and converge to the steady state faster than the existing technique of observing channel busy status alone [136].

5.2 Markov model of the Broadcasting DCF

By modelling the broadcasting non-QoS-enabled IEEE 802.11 DCF backoff counter using a variant of discrete time Markov model presented by Bianchi *et al.* [148], adapted for broadcast transmissions, the behaviour of it can be analysed. The symbols used in this model are listed in Table 5.1.

This model uses only the top row of states in the Bianchi model, and discards the remaining states representing the exponential backoff procedure. Similar to the original model, this model quantises time into "slots" of varying lengths,

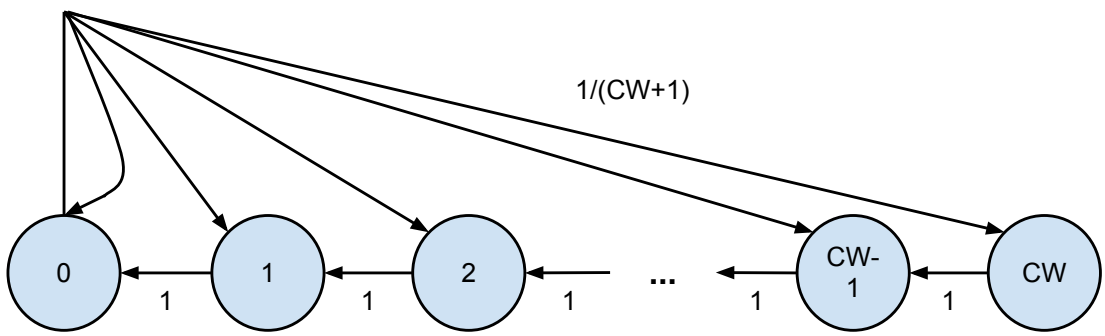


Figure 5.1: Markov chain model for a single saturated station

delineated by the decrement of the backoff counter. Stations are assumed to be saturated (always have something to send) and are synchronised. Since this analysis is not concerned with throughput or other time-related measures, the length of each slot is unimportant. Each state in the model is represented by the value of the backoff counter at that state. Figure 5.1 depicts this model graphically. If a backoff counter has a value of $X \in [1, aCW_{min}]$, it will always have a value of $X - 1$ at the next slot. If a backoff counter has a value of 0, the frame will be transmitted and the counter reset to a value uniformly distributed in the contention window $[0, aCW_{min}]$ as per IEEE 802.11 specifications.

5.2.1 Analysis of steady state probability

Based on the Markov model, the steady state probability of being in any state can be defined recursively as:

$$P'_k = \begin{cases} P_{k+1} + \frac{P_0}{CW+1} & k \in [0, CW) \\ \frac{P_0}{CW+1} & k = CW \text{ where } CW = aCW_{min} \end{cases} \tag{5.1}$$

Given that the sum of all P_k equals to 1:

$$\begin{aligned} 1 &= P_0 + P_1 + P_2 + ... + P_{CW-1} + P_{CW} \\ &= P_0 + P_1 + P_2 + ... + (P_{CW} + \frac{P_0}{CW+1}) + \frac{P0}{CW+1} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{CW+1} \frac{kP_0}{CW+1} \\
&= \frac{P_0}{CW+1} \sum_{k=1}^{CW+1} k \\
&= \frac{(CW+1)(CW+2)}{2} \frac{P_0}{CW+1} \\
&= \frac{P_0(CW+2)}{2} \\
P_0 &= \frac{2}{CW+2}
\end{aligned} \tag{5.2}
$$

Hence:

$$
\begin{aligned}
P_k &= (CW+1-k)\frac{P_0}{CW+1} \\
&= \frac{2(CW+1-k)}{(CW+1)(CW+2)}
\end{aligned} \tag{5.3}
$$

This formula represents the overall probability of a station's backoff counter having a value of k.

Now, unlike in Bianchi and Tinnirello's approach [136] where they solved the expression for collision probability, this analysis attempts to determine the number of backoff slots between transmissions.

5.3 Relationship between contention and inter-frame slots

5.3.1 Naïve solution based on binomial expansion

A naïve solution using the Markov model is to put the steady state probabilities into a simple binomial expansion. Assuming there are N saturated stations, $aCW_{min} = CW$:

For slot $k = 0$:

The probability of not having any transmissions can be written as:

$$
\begin{aligned}
{}^{0}T_0 &= \binom{N}{0}(1-P_0)^N \\
&= (1-P_0)^N
\end{aligned}
\tag{5.4}
$$

Hence the probability of one or more transmissions is:

$$
\begin{aligned}
T_0 &= 1 - {}^{0}T_0 \\
&= 1 - (1-P_0)^N
\end{aligned}
\tag{5.5}
$$

Assuming collision is the only cause of frame loss, and any collision will cause all colliding frames to be lost, the probability of successful Rx is:

$$
\begin{aligned}
R_0 &= \binom{N}{1}(P_0)^1(1-P_0)^{N-1} \\
&= NP_0(1-P_0)^{N-1}
\end{aligned}
\tag{5.6}
$$

and the probability of collision:

$$
X_0 = T_0 - R_0 \tag{5.7}
$$

For slot $k \in [1, aCW_{min}]$:

Here, only the conditional probability given there had not been any transmissions earlier in the contention window needs to be considered. If there had been prior transmissions, the procedure would have been reset (*i.e.* $P_{k'|T_k} = 0 \quad \forall k \in [0, k')$). This conditional probability can be written as:

$$
\begin{aligned}
{}^{0}T_{k|{}^{0}T_{k-1}} &= \binom{N}{0}\left(1-\frac{P_k}{\sum_{m=k}^{CW} P_m}\right)^{N} \\
&= \left(1-\frac{P_k}{\sum_{m=k}^{CW} P_m}\right)^{N}
\end{aligned} \tag{5.8}
$$

Therefore the probability of not having any transmission since the start of the contention window is:

$$
\begin{aligned}
{}^{0}T_{k} &= {}^{0}T_{k|{}^{0}T_{k-1}} \times {}^{0}T_{k-1} \\
&= {}^{0}T_{k-1}\left(1-\frac{P_k}{\sum_{m=k}^{CW} P_m}\right)^{N}
\end{aligned} \tag{5.9}
$$

and the probability of the first transmission being at slot k is:

$$
\begin{aligned}
T_k &= (1-{}^{0}T_{k|{}^{0}T_{k-1}}) \times {}^{0}T_{k-1} \\
&= {}^{0}T_{k-1}\left(1-\left(1-\frac{P_k}{\sum_{m=k}^{CW} P_m}\right)^{N}\right)
\end{aligned} \tag{5.10}
$$

The probability of the first transmission being at slot k and is successful is therefore:

$$
\begin{aligned}
R_k &= \binom{N}{1}\left(\frac{P_k}{\sum_{m=k}^{CW} P_m}\right)^{1}\left(1-\frac{P_k}{\sum_{m=k}^{CW} P_m}\right)^{N-1} {}^{0}T_{k-1} \\
&= N\,{}^{0}T_{k-1}\left(\frac{P_k}{\sum_{m=k}^{CW} P_m}\right)\left(1-\frac{P_k}{\sum_{m=k}^{CW} P_m}\right)^{N-1}
\end{aligned} \tag{5.11}
$$

and the probability of the first transmission being at slot k and is unsuccessful can be expressed as:

$$X_k = T_k - R_k \tag{5.12}$$

Overall statistics:

The probability of successful transmission is:

$$R = \sum_{k=0}^{CW} R_k \tag{5.13}$$

The probability of collision is:

$$\begin{aligned} X &= \sum_{k=0}^{CW} X_k \\ &= 1 - R \end{aligned} \tag{5.14}$$

And the expected number of slot between transmissions is:

$$E[T] = \sum_{k=0}^{CW} kT_k \tag{5.15}$$

5.3.2 Accounting for observation dependencies

The naïve solution gives the steady state probability, assuming that system observations are taken in a process that is independent from the underlying states. When observations can only occur at specific states (*i.e.* when a station transmits after reaching state 0), some stations may not have reached steady state and therefore this naïve model may not fit well. This is especially evident when the number of stations (N) is either too small or too large compared to the contention window size. After transmitting a frame, a station reinitialises its backoff counter to a uniformly distributed value (thus does not follow the stationary

probabilities of the Markov model). The ideal model would account for the entire history of each station (breaking the Markov model assumption), and would be intractable for most cases. As a compromise, this thesis accounts for the number of stations last transmitted and uses uniform distribution instead of the steady state probabilities of the Markov model for these stations, improving the model by considering the one-step history at each station.

To allow for multiple concurrent transmissions, one needs to both incorporate the various expressions for concurrent transmissions and to determine the likelihood of its occurrence.

In this analysis the uniform distribution over the contention window $[0, CW] \subset \mathbb{Z}$ is denoted as U, and the probability of choosing slot k from this distribution $U_k = \frac{1}{CW+1}$.

The probability of not having any transmissions at slot k, given there were no transmissions in the previous slot, is:

$$ {}^{0}T_{k|{}^{0}T_{k-1}} = \sum_{i=1}^{N} {}^{i}T \binom{i}{0} \left(1 - \frac{U_k}{1 - \sum_{j=0}^{k-1} U_j}\right)^{i} \binom{N-i}{0} \left(1 - \frac{P_k}{1 - \sum_{j=0}^{k-1} P_j}\right)^{N-i} \tag{5.16} $$

Here, i is the number of concurrent transmissions in the last cycle.

To compute the probability of having n concurrent transmissions at slot k given there was no transmission in the previous slots, all possible ways the n stations could be distributed between the set of previously transmitted stations (which follows uniform distribution) and the set of stations that did not transmit in the last cycle (and follows the Markov chain) need to be considered:

$$ {}^{n}T_{k|{}^{0}T_{k-1}} = \sum_{i=1}^{N} {}^{i}T \sum_{m=\max(0,n-N+i)}^{\min(i,n)} \binom{i}{m} \left(\frac{U_k}{1 - \sum_{j=0}^{k-1} U_j}\right)^{m} \left(1 - \frac{U_k}{1 - \sum_{j=0}^{k-1} U_j}\right)^{i-m} $$

$$\binom{N-i}{n-m}\left(\frac{P_k}{1-\sum_{j=0}^{k-1}P_j}\right)^{n-m}\left(1-\frac{P_k}{1-\sum_{j=0}^{k-1}P_j}\right)^{N-i-n+m} \tag{5.17}$$

Using these, the unconditional probabilities ${}^{n}T_k$ can be determined as:

$${}^{n}T_k = {}^{n}T_{k|{}^{0}T_{k-1}} \prod_{j=0}^{k-1} {}^{0}T_{j|{}^{0}T_{j-1}} \quad \text{where } {}^{0}T_{-1} = 1 \tag{5.18}$$

Finally, the probability of n concurrent transmissions can be determined by summing over all slots:

$${}^{n}T = \sum_{k=0}^{CW} {}^{n}T_k \tag{5.19}$$

5.3.3 Numeric solution to the system of equations

Computing the probability of concurrent transmissions (${}^{n}T$ for some $n \in [1, N]$) requires solving a system of polynomials of degree CW with N variables. (The steady state ${}^{n}T$ is a polynomial of ${}^{j}T \quad \forall j \in [1, N]$ of degree CW.) While it is possible to determine exact solutions of such systems using techniques such as computing the Gröbner basis, algorithms to find these basis are complex. The most commonly used algorithm, implemented in Matlab, Mathematica and other software, is the Buchberger's Algorithm [182]. Extending Mayr's results [183], it can be shown that using this algorithm to solve the system has a worst case complexity of $O(CW^{2^N})$, and is therefore not viable for our analysis. Therefore, numeric approximations are used to solve the system of equations representing the analytic solution.

To approximate the various ${}^{n}T$, an initial approximate is calculated by dropping all terms with degree greater than 1 and solving the resultant set of linear equations. The approximate is then improved incrementally using an adapted form of binary search where, for each ${}^{n}T$, the mean between the approximate

Algorithm 5.1 Numeric solution of concurrent transmission probabilities

```
function COMPUTE_APPROXIMATE(S: set of polynomial equations for ^nT)
    S' ← drop terms of degree > 1 for each equation in S
    A ← solution of S' using Gauss-Jordan elimination
    while true do
        A ← normalise A such that     A = 1
        A' ← substitute A into S and evaluate
        if max(A') < 5 × 10^-4 or max(|a − a'|) < 10^-9   ∀a ∈ A, a' ∈ A' then
            return current best estimate A
        end if
        if more than 20 retries then
            return current best estimate A
        end if
        if first retry then
            A ← mean of A and A'
            continue
        end if
        if odd retries then
            A ← move A' the opposite direction to last perturbation
        else if even retries then
            Flatten the vector A — reduce peak by 3% and distribute uniformly
to the remainder
        end if
    end while
end function
```

and the result are first normalised to 1 and then used as the initial approximate to the next increment. If the mean does not improve the estimate, the guess is then slightly perturbed either side of the guess and/or the result vector flattened for up to 20 times, retrying the new guess afterwards. The incremental step is iterated until an error of less than 5×10^{-4} is achieved or no improvements can be made. This algorithm is presented as Algorithm 5.1.

Algorithm 5.1 was implemented in C++ and run on a computer with one Intel E8400 CPU and 4 GiB of RAM. Analytic results are computed for up to 450 stations and contention window up to 255. Execution of the approximations run for no more than 3 days.

5.4 Accuracy of DCF model

Computer simulations were conducted using a simplified model of the DCF to verify the correctness of the model. Complex simulation packages such as ns-3 were not used in the investigation because these packages also simulate more complex physical layer effects that complicate the interpretation of results. Simulations are conducted for CW sizes of 3, 7, 15, 31, 63, 127 and 255, with up to 100 saturated stations within range of each other. The statistics on collision probability and interframe idle slot counts are collected and then compared to the model predictions.

5.4.1 Simulated DCF model

A model simulating a DCF backoff counter was constructed to verify the mathematical model. This simulation model is a simple decrementing counter that is reset once it reaches zero, and assumes perfect physical channel. The model simulates the following:

- Fixed size contention window (CW) for each station;
- Backoff counter reinitialise to a uniformly distributed value within the CW after transmission by the station. This models the DCF broadcast behaviour (*i.e.* no ACKs) and assumes all stations are saturated;
- Global (shared) timeline in "slots". Data transmission, IFS, *etc.* occur between slots and the actual wall time for the action is ignored;
- Transmission is lost if and only if there is a collision (two or more stations scheduled to transmit in the same slot); and
- The model assumes all stations are synchronised (propagation and processing times are zero and no hidden stations). Without assuming synchronisation, the time between slots cannot be ignored as stations that are not synchronised will see different slot boundaries.

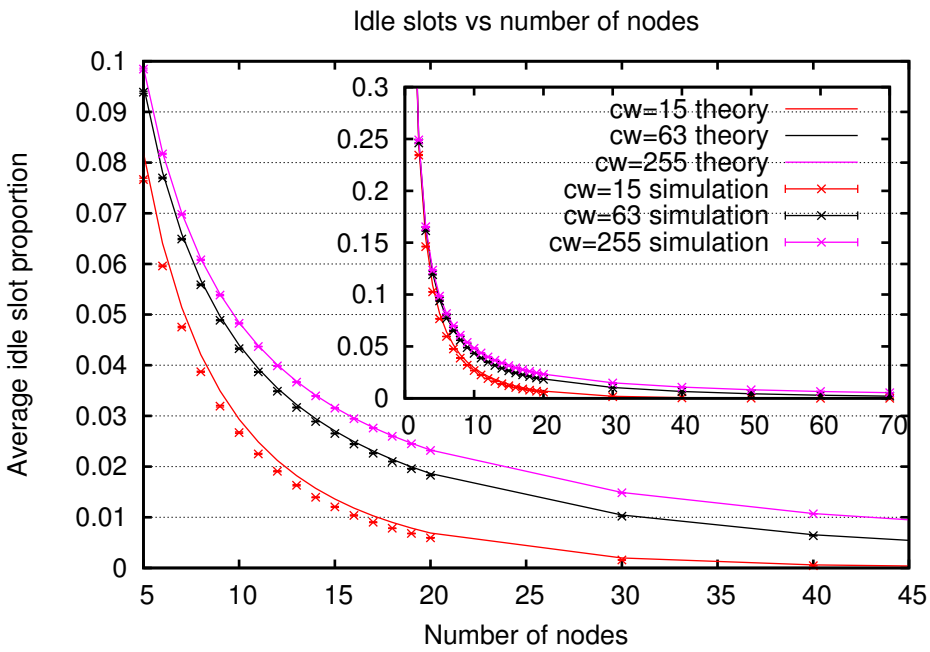
Each station simulated is initially assigned a backoff counter value uniformly distributed over the contention window. At each time step, all backoff counters are decremented by one if the counter value is greater than zero. If the counter is zero, it is assumed that the station will initiate a transmission, and the counter is reset to a backoff counter uniformly distributed over the contention window. The transmission is assumed to be successful if only one station initiated a transmission, and assumed to have failed due to collision if more than one station transmitted. If no station initiated a transmission at that timeslot, then the channel is considered idle at that time, otherwise, the channel is considered busy. In this simulation, statistics on idle periods, probability of channel being busy and packet success ratio are collected.

This model uses the Combined Multiple-Recursive Generator MRG32k3a [184] as the pseudo-random number generator. This generator is also used in many complex network simulation packages such as ns-2 and ns-3. This pseudo-random number generator has a period of 2^{191}, and the implementation used in the simulator divides this into 2^{64} non-overlapping subsequences of 2^{127}. Each execution of the model uses a different seed, thus has a good probability that the executions are statistically independent from each other. In the experiments, a sample of 500,000 idle periods was collected for each station count–CW pair.

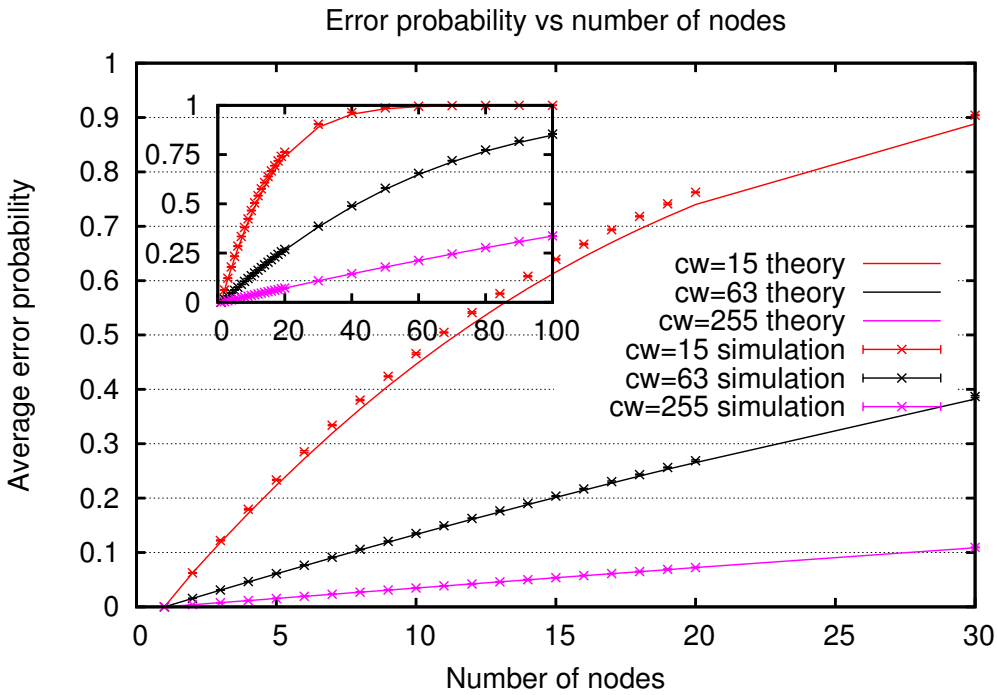
5.4.2 Results

Figure 5.2 compares the overall network statistics between the model prediction and the simulation results using the simple DCF model. Further comparison looking at the distribution of backoff slot for a contention window of 63 and varying number of stations are included in Figure 5.3.

The simulation results (Figure 5.2) suggest that the expected number of idle slots decreases and the collision probability increases as the number of saturated stations increases. This confirms the intuition that as more stations try to transmit, the chance that some station would transmit while another is still decre-



(a) Mean idle slots proportion



(b) Mean proportion of packet failures due to collisions

Figure 5.2: Overall network statistics for a contention window size of 64 as a function of total number of concurrent saturated stations, in the absence of hidden stations, as predicted by the theoretical model vs simulation results. Simulation results are aggregated over 10 executions of the simulation using different random seeds. Error bars denotes two standard deviations from sample means.

Table 5.2: Goodness-of-fit — Overall network statistics

CW	15		63		255	
	Idle Slots	Error	Idle Slots	Error	Idle Slots	Error
R^2	0.99893	0.99755	0.99993	0.99986	0.99999	0.99997

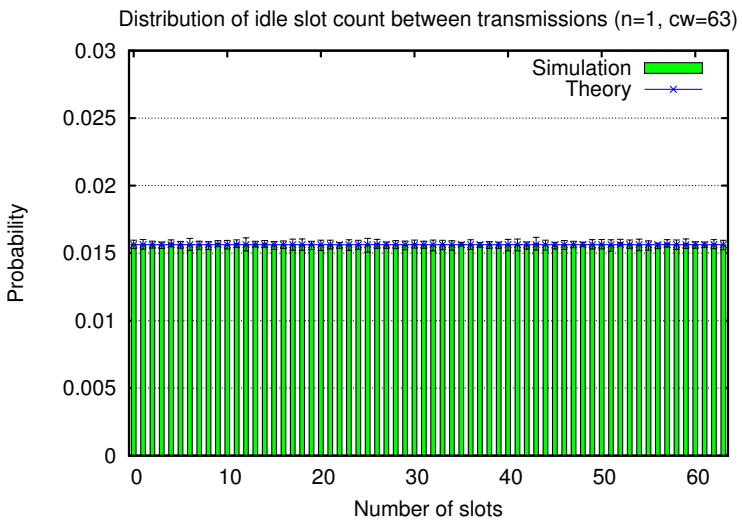
Table 5.3: Goodness-of-fit — Predicted distribution ($aCW_{min} = 64$)

#Stations	1	5	10	15	50	150
R^2	1.4e-11	0.9998	0.9999	0.9999	0.9999	1.0000

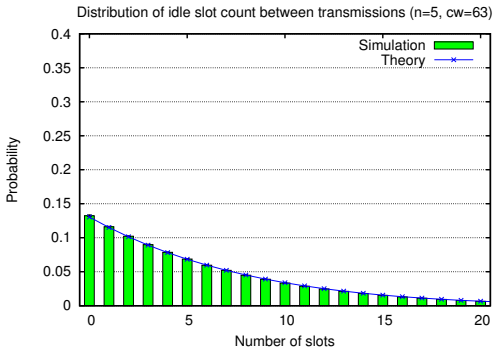
menting the backoff counter increases. This also confirms the intuition that when the number of stations increases, the chance of two of more stations choosing the same backoff slot also increases.

By comparing the overall statistics from the theoretical prediction to the simulation output in Figure 5.2, the accuracy of the theoretical model in predicting the expected idle slot counts and the associated error probabilities can be confirmed graphically. In addition, Table 5.2 calculates the R^2 value for the model. Both the idle slots and the success/error predictions have R^2 close to 1, indicating very high correlation between the observed data and the model.

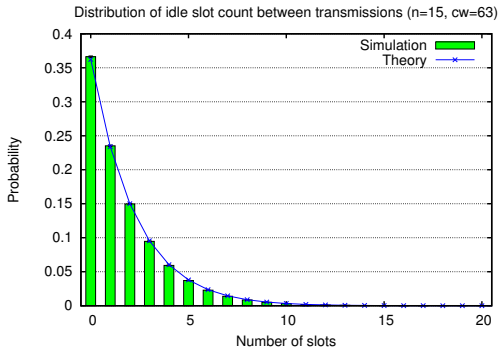
Figure 5.3 further compares the performance of the model with the simulation, looking at the probability distributions when the contention window size is restricted to 64. In general, these plots indicate that the theoretical model presented is quite accurate in predicting the probability distribution of idle slot counts. Figure 5.3d and e both showed that the theoretical model very slightly underestimates the probability of immediate transmissions at very high station densities (50 and 150 stations in range). This small discrepancy would explain the underestimation of packet loss observable in Figure 5.2b. Table 5.3 shows the R^2 values for these predictions. All results except for $N = 1$ shows a very high R^2 value, indicating high correlation between the prediction and the observations. For the case $N = 1$, since the prediction is a horizontal line, the R^2 value cannot provide a useful measure of correlation. Nevertheless, the good fit between the model predictions and the data can be confirmed visually using graphical means.



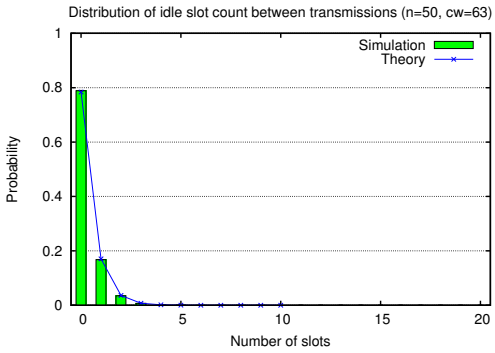
(a) 1 station (y-axis range from 0 to 0.03 only). ($\frac{1}{64} = 0.015625$)



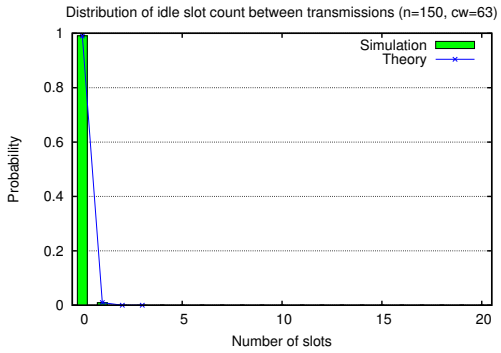
(b) 5 stations



(c) 15 stations



(d) 50 stations



(e) 150 stations

Figure 5.3: Distribution of idle slots between transmissions (interframe space) for a contention window size of 64, as predicted by the model (line) vs simulation results (columns) using the simplified DCF model. Simulation results are aggregated over 10 executions of the simulation using different random seeds. Error bars show two standard deviations from the sample means, most are too small to be visible.

5.4.3 Discussion

The investigation in this section raises concerns regarding the capacity of the current IEEE 802.11p standard for vehicular safety systems. The expected capacity of the current IEEE 802.11p configuration (where aCW_{min} is 15) is obvious from the $CW = 15$ plot in Figure 5.2b. Current proposals for vehicular safety applications require a minimum packet reception ratio of 90%. However, even two saturated stations on the network will degrade the PRR to the threshold without considering physical effects that contributes to frame loss! It must however be noted that this simulation does not consider the effect of distance and packet capture, which can enhance reception, nor does it consider any "link layer desynchronisation" [185] where the MAC timers are not synchronised across the network (typically caused by hidden terminals, shadowing and fast fading). Torrent-Moreno *et al.* found that stations as close as 47% of the transmission range (and 24% of carrier sense range) may not be able to detect a concurrent transmission due to fast fading alone [185]. The lack of synchronisation between stations may cause more collision than is predicted by the model. While it is noted that saturated stations may be highly unrealistic in practice, the work in the upcoming chapters will relax this condition and show this capacity limitation to apply even for unsaturated stations.

The investigation in this section therefore suggests that, in order for vehicle-to-vehicle communication to meet the target PRR, load offered to the channel must be tightly controlled. Enabling the Medium Access Control function to sense and adapt to channel load using implicit feedback mechanisms such as passive channel observations may be necessary.

5.5 Estimating channel load

Having demonstrated that the model can predict the channel behaviour for a given number of concurrent saturated stations on the network, this theoretical

Algorithm 5.2 Algorithm to measure the number of saturated stations

Let $\boldsymbol{b}$ be the belief vector of saturated stations count.
for all $b_i \in \boldsymbol{b}$ **do**
  $b_i \leftarrow \frac{1}{\text{length}(\boldsymbol{b})}$
end for

loop
  $\hat{o} \leftarrow$ observed number of idle slots
  $denominator \leftarrow 0$
  for all $b_i \in \boldsymbol{b}$ **do**
    $denominator \leftarrow denominator + b_i(T_{\hat{o}}|N = i)$
  end for
  for all $b_i \in \boldsymbol{b}$ **do**
    $b'_i \leftarrow \frac{b_i(T_{\hat{o}}|N=i)}{denominator}$
  end for
  $\boldsymbol{b} \leftarrow \{b'_i \quad \forall i\}$
end loop

$N_{est} \leftarrow \sum_i ib_i$ ▷ Estimated station count is the weighted sum of belief

model is then used to estimate the number of concurrent transmitting saturated stations by observing the distribution of interframe spaces.

In this section, a channel load estimator that uses Bayesian inference to estimate the most likely number of saturated stations on the network is described and evaluated. Bayesian inference is based on Bayes theorem in probability theory such that observed outcomes are used to derive a distribution of the underlying factors on which the observations are conditional upon. It is simple to implement, and are quite accurate in practice. Algorithm 5.2 outlines the operation of the load estimator.

The resultant belief vector $\boldsymbol{b}$ represents the likelihood that the current estimate is the correct number of saturated stations on the network. One method to interpret this belief vector is by taking the entry with the highest probability (Maximum Likelihood). However, since the number of categories used in this estimator is much smaller than the domain of the conditions, the weighted sum of the belief vector is taken as the estimated contention value. This allows the estimator to interpolate for the number of saturated stations that is not in the referenced set.

5.5.1 Performance evaluation

In this experiment, a network containing a fixed number of saturated stations is simulated, each operating as described in Section 5.4.1. Additionally, one passive observer station is inserted into the network, and is configured with two channel load estimators. During the simulation, channel observations are fed to both channel estimators on this station, with the estimates from these estimators compared to the parameters of the simulation. The experiment compares the estimation by observing interframe spaces with Bianchi and Tinnirello's approach [136] of observing channel busy status at each time slot in terms of both accuracy and the time to reach steady state.

In this experiment, one of the channel estimators is configured with the probability distribution of idle slots calculated by the theoretical model, and are given the idle slot observations every time a transmission occurs. The other estimator is configured with the steady state probability of transmission in any given time slot. This method adapts Bianchi and Tinnirello's approach [136] but removes from their Markov model all the states related to MAC retransmissions, which are not experienced in a broadcast environment. The estimator configured to observe channel busy status is fed channel observation (busy or not) at every time slot.

It is noted that, unlike in Bianchi and Tinnirello's paper [136], simple Bayesian inference is used. Not only this technique very easy to implement, it also provides a common basis to compare the two approaches.

For ease of implementation in this experiment, the number of concurrent saturated stations is limited to a finite set of discrete integers. Even though the function that calculates the expected idle slot distribution for any given number of saturated stations has a domain that spans the entire set of positive integers, the calculation of the actual values are not easily performed. Hence only the probability distributions of idle slots for the small subset of saturated station counts are precomputed. The set of idle slot distributions are chosen to be the data points originally obtained for the previous section (Figure 5.3), (including those

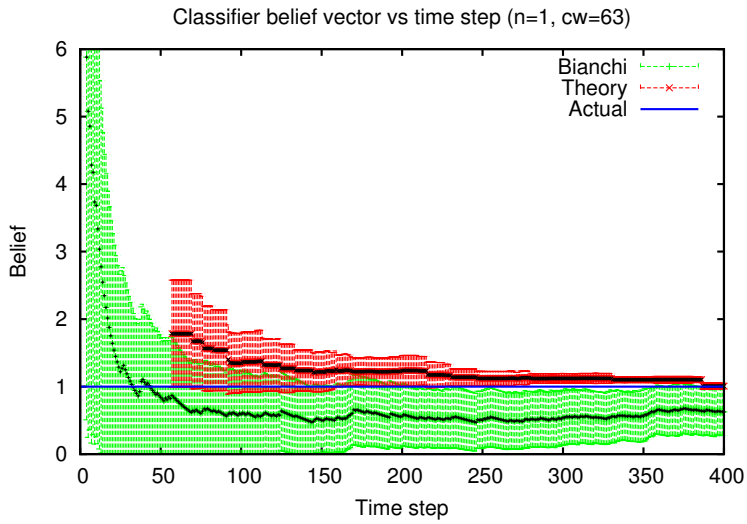
collected but not plotted in the figure). The data points are chosen mainly due to convenience and are not the category that optimises for accuracy or usefulness in prediction outcome.

5.5.2 Results

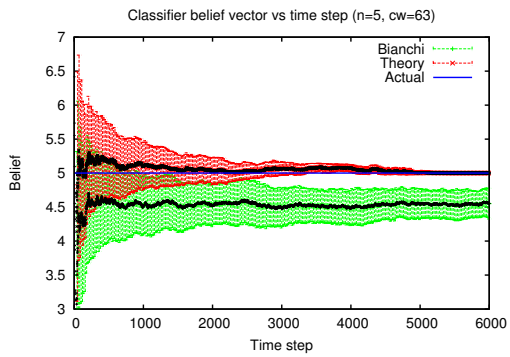
Figure 5.4 plots the respective belief vectors from the two channel load estimators. It should also be noted that Bianchi and Tinnirello's approach assumes the observing station is always transmitting (saturated), thus the actual output from this estimator is one higher than the number of saturated stations. Figures 5.4 and 5.6 have been adjusted to account for this behaviour.

Overall, the channel load estimator observing idle slot counts outperforms the one observing the channel busy status ("collision probability") in terms of both the estimation accuracy and the time to steady state. Figure 5.4 shows the mean and standard deviation of the probability distribution in the belief vector over time for a system with contention window size of 64 ($aCW_{min} = 63$). The number of saturated stations tested include the case with only a single station (Figure 5.4a), less stations than number of slots available (Figures 5.4b and c), number of stations close to number of slots available (Figure 5.4d) and the number of stations exceeds the number of backoff slots available (Figure 5.4e).

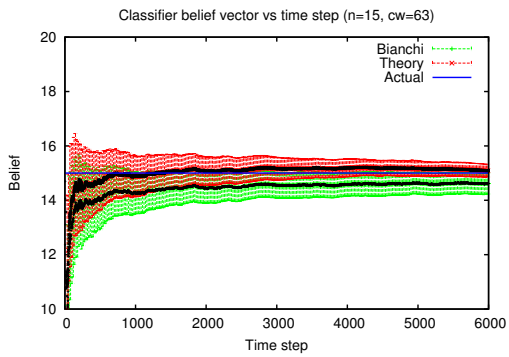
Figure 5.4 shows that both Bianchi and Tinnirello's [136] and our model can be used with Bayesian inference to determine the number of saturated stations based on channel observations. The black points mark the means of the belief vectors, and the error bars show the spread of the probabilities (one standard deviation). The red error bars are the belief vectors from the estimator observing idle slot counts, whereas the green error bars correspond to the estimator observing channel busy status. These figures show that as time progresses, the means of the belief vectors for both techniques converge to a value close to the actual parameter (*i.e.* the estimates are accurate), and the spread of the probabilities reduces (*i.e.* the estimator is becoming more certain).



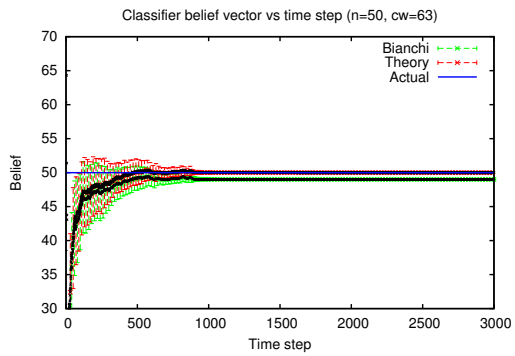
(a) 1 saturated station (x-axis range up to 400 steps only).



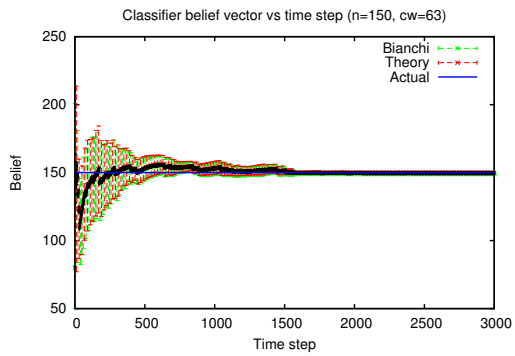
(b) 5 stations



(c) 15 stations



(d) 50 stations



(e) 150 stations

Figure 5.4: Estimator belief vector as time progresses for a contention window size of 64. Plotted is the mean and standard variation of the belief vector probability distribution. The blue solid line represents the true configuration of the simulation. Only one in 20 data points are plotted to allow the other line to show through. Early values for the red "theory" points (from the idle slot observation method) are outside the plot until the after the first backoff period — *i.e.* the first belief vector update.

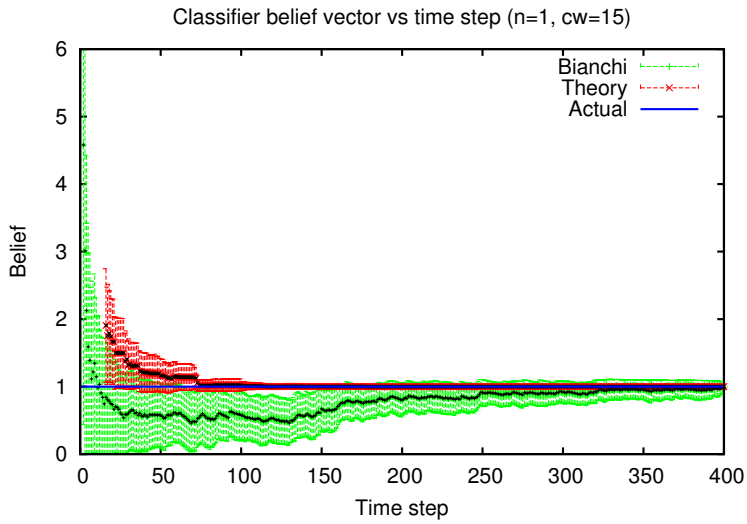
It is noted that the early data points for the estimator observing idle slot counts are outside the range of the y-axis in Figure 5.4a. This is due to the fact that, unlike the Bianchi technique, the belief vector is not updated until the first transmission, and it may take a few observations before the belief vector gets within the range of the y-axis.

In the scenarios tested and shown in Figure 5.4, the estimator that observes idle slots tends to converge to a value closer to the true simulation parameter. The mean estimate from Bianchi and Tinnirello's technique tends to be half to one station lower than the actual parameter. In addition, the estimator observing channel busy status is slower to become certain about the estimate, as can be seen by the higher spread.

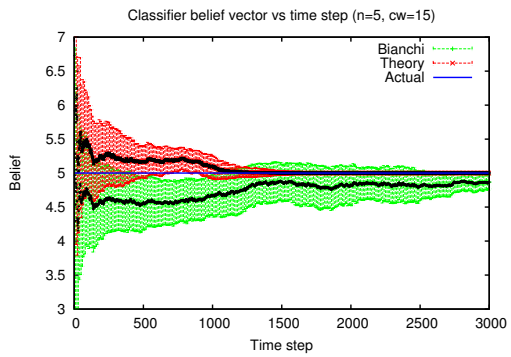
The outcomes when the contention windows is only 16 slots long show similar trends. Figure 5.5 shows that both techniques converge to an estimate close to the actual parameter value, and their confidence in their estimates also grow (reducing probability spread) as time progresses. In these tests, Bianchi and Tinnirello's technique still seems to underestimate at low load, but on the other hand the presented technique tends to overestimate at high load.

Estimating non-referenced values

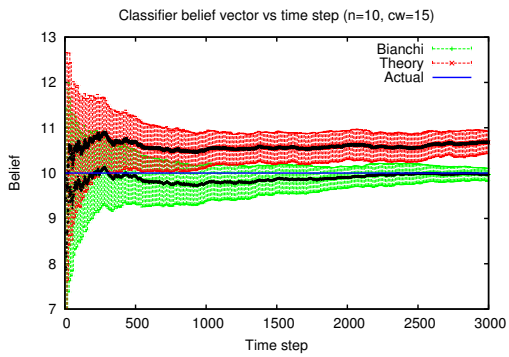
When the number of saturated stations are not in the set of reference values, the estimators are likely to eventually choose as result a member of the reference set instead of the true value. Figure 5.6 compares the belief vectors between (a) a non-referenced number of saturated stations (24, the closest categories are 20 and 30), and (b) 19 saturated stations, which is an element of the reference set. The scenario with 20 saturated stations cannot be used here due to assumption in Bianchi and Tinnirello's approach that the observer station is also saturated, giving an expected output of 21, which is not within the reference set. In these figures, the reference values common to both estimators are coloured black, ones unique to Bianchi and Tinnirello's technique (due to the off-by-one behaviour)



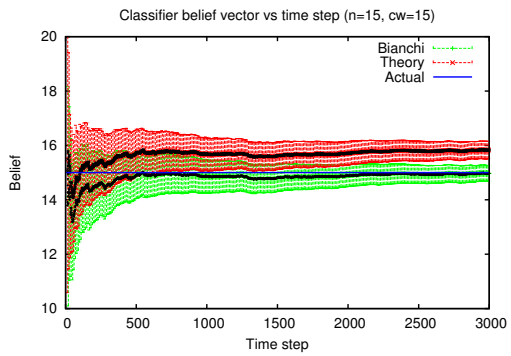
(a) 1 saturated station (x-axis range up to 400 steps only).



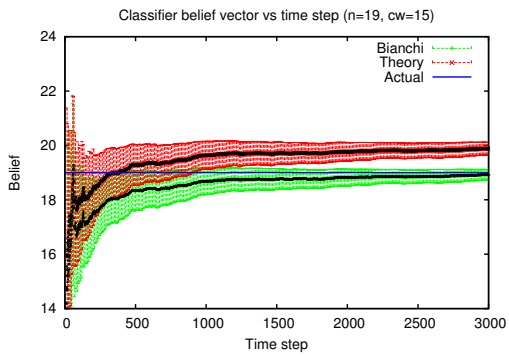
(b) 5 stations



(c) 10 stations

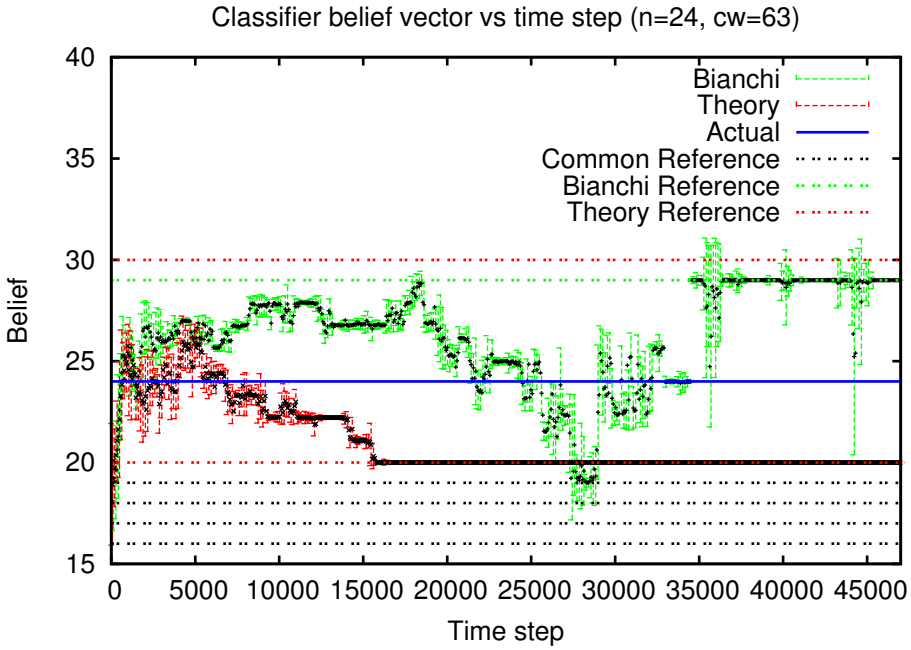


(d) 15 stations

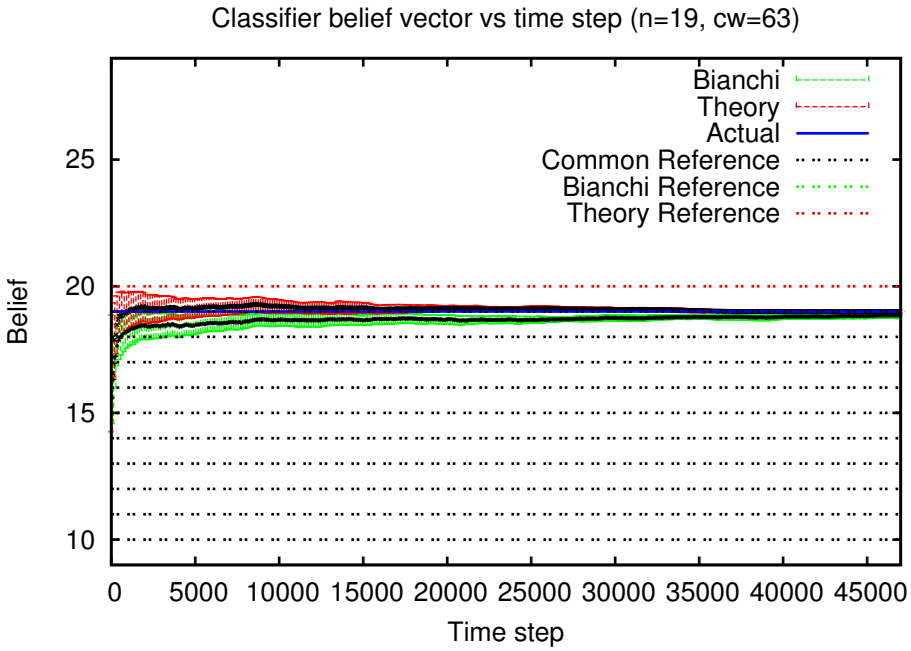


(e) 19 stations

Figure 5.5: Estimator belief vector as time progresses for a contention window size of 16. Plotted is the mean and standard deviation of the belief vector probability distribution. The blue solid line represents the true configuration of the simulation. Only one in ten data points are plotted to allow the other line to show through. Early values for the red "theory" points (from the idle slot observation technique) are off the plot until the after the first backoff period — *i.e.* the first belief vector update.



(a) 24 saturated stations — not in reference set



(b) 19 stations — an element of the reference set

Figure 5.6: Comparison of estimated contention levels when the actual number of saturated stations is either (a) outside the reference set, or (b) an element of the reference set. Estimated contention level is plotted as the weighted sum and the standard deviation of the estimator belief vectors. Contention window size of 64 was used. Only one in 70 points are plotted to allow other error bars to be visible. Plot shows a much noisier output when contention level is not within the reference set, with the output converging to a value in the reference set instead of the true value.

are green, and the ones unique to the idle slots approach are coloured red.

Figure 5.6 shows that when the number of saturated stations is not within the reference set, the channel load estimator output may eventually converge to a value within the reference set instead of the true value, with the output from the estimators being much noisier. Consistent with earlier observations, the estimates for the 19-station test (Figure 5.6b) converges consistently to the correct value for both estimators. In comparison, the 24-station tests (Figure 5.6a) converge much slower to their steady states, with the mean lingering around the true value for an extended period of time before stabilising at a referenced value. The estimator observing channel busy status shows occasional large variances in its estimates even after the mean value reached its steady state.

5.5.3 Discussion and future work

It can be seen that the output from Bianchi and Tinnirello's technique is noisier across the tests conducted. The noise is more prominent with smaller contention window sizes probably because the technique only uses the binary busy/idle status to estimate channel condition. This means individual states may be much more influential on the mean. It also explains why interframe period observations generate a less noisy result. The wider range of possible outcomes from observing idle slots is also beneficial for improving confidence in the estimates, thus allows the estimates to converge faster despite the lower update frequency.

This work also highlights the need for appropriate windowing strategies or the use of more sophisticated classification/regression algorithms. Simple Bayesian inference retains infinite history, therefore it cannot track changing channel conditions. Retaining infinite history means that when the number of samples is large enough, additional samples would provide minimal influence on the estimator unless an extremely rare event is observed. Figure 5.6a showed that as time progresses, the estimator output converges to one of the reference values due to the lack of windowing strategy. Experimental results suggest that the estima-

tor output may linger near the true value for some time before converging to the steady state value. Hence an appropriately chosen aggregation window (size and/or shape) could potentially avoid the estimator getting stuck at a certain category, and allows the estimator to track changing conditions.

The use of simple Bayesian inference in this section is only intended to be a sample application of the model. The theoretical model presents a method to compute the number of idle slots that can be expected for any number of saturated stations in the network. This work also demonstrated that one may use observed interframe periods to estimate the channel load in terms of the number of saturated stations. In actual practice, one can use the presented model and the technique of observing idle slots in estimators other than (or in conjunction with) simple Bayesian inference to improve the accuracy and/or time to steady state. One may, for example, apply a hamming window, ARMA and/or EKF filter [136], MAP filter [137], or Viterbi algorithm over the output of the Bayesian estimator [138]; or to use particle filter techniques in place of the Bayesian estimator similar to [139]. The work presented in this section can be used as the basis of any applicable classification/regression techniques in order to estimate channel load.

Furthermore, when compared to Bianchi and Tinnirello's technique [136] of observing frame collisions, observing idle slot counts converges faster despite the lower refresh rate. Observing channel busy status causes the estimator to slowly adjust its belief vector at every slot, whereas observing idle slots cause large adjustments every few slots. Since channel contention is unlikely to change much between slots, the lower refresh rate does not affect the usefulness of the technique. On the other hand, the faster convergence enables the use of smaller aggregation windows, thereby allowing estimators that do not retain full history to track current channel contention quicker in a dynamic environment.

Finally, in modelling this system, all stations are assumed to be within range of all others and are saturated. This is atypical in real life. When a station cannot

sense a current transmission, the assumption that all stations are synchronised becomes invalid, thus may invalidate the model. Intuitively, hidden terminals might transmit during another station's transmission, thus one may no longer disregard the timing aspect of the scenario, and cannot use flexible slots as the unit of time (without assuming transmission takes integer number of slots). Additionally the state transition of one station is no longer independent of another station. For these reasons, a model that allows hidden terminals cannot be Markovian in the current form. Bastani *et al.* used a slight variant of this Markov model whereby time is still quantised into slots, but the DCF state may not update at each slot depending on the channel condition [51]. A similar extension of the model presented may be useful for incorporating hidden terminals. Further investigations on the actual effect of both unsaturated stations and hidden terminals on the idle slot distribution is needed.

5.6 Conclusion

In this chapter, the Bianchi model was used to derive an expression relating the number of idle slots between IEEE 802.11broadcast transmissions, to the number of saturated stations on the network. A channel contention measurement technique exploiting this relationship was described. The described technique uses simple Bayesian inference and observes idle slot counts between frames. Through computer simulations, it is shown that this technique is effective in estimating the number of saturated stations on a network with no hidden terminals. When compared to the existing technique of observing packet collision probability, the technique of observing idle slot counts reaches steady state faster, with the estimate being closer to the true value.

Furthermore, investigations in this chapter revealed a potential issue with channel capacity for vehicular networks — the IEEE 802.11p channel would degrade to below the required 90% reception with only two saturated stations on the network. It is identified that appropriate MAC-layer channel contention sensing

with appropriate windowing mechanisms will be necessary for controlling channel contention and thereby improving packet reception ratio.

In the next chapter, the effects of unsaturated stations on the channel and the channel estimation algorithms are evaluated, leading to the development of an extension to the broadcasting DCF model.

Overview

This chapter investigates the applicability of the passive channel observation technique to networks with stations that are not saturated. The measure "Equivalent Saturated Node" (ESN) is introduced both to describe the level of saturation of a station, as well as the level of contention across a network. An extended theoretical model to analyse unsaturated stations is introduced and is found to be not viable.

Contributions

- **I demonstrated the effects unsaturated stations have on the relationship between the wireless channel contention and the observed idle slot counts and their impacts on channel contention estimation techniques.** Here, channel contention is defined as the sum of individual stations' saturation across all stations in the network. Through simulation, I showed that station saturation has a small but observable effect on both the distribution of idle slot count and the collision probability. There are minor impacts on the estimators' channel contention estimation accuracy as well as slight lengthening of time before the estimates stabilise. I have also shown that the technique of observing idle slot counts is more resilient to errors caused by unsaturated stations.

Chapter 6

Networks with Unsaturated Stations

6.1 Introduction

In the previous chapter, networks containing only saturated stations were investigated, whereas real networks are unlikely to contain many of them (if at all). This chapter investigates the applicability of those results to the more generic situation with unsaturated stations. In particular, a measure called "Equivalent Saturated Nodes" (ESN) is introduced to describe both the level of contention in a network as well as the degree of saturation of individual stations. Using this measure, the DCF model presented in Chapter 5 and the passive channel load estimator is tested against various station saturation levels to determine their applicability in situations with homogeneous unsaturated stations.

An extension to the DCF model incorporating unsaturated stations is also presented, with its predictions tested against simulation outcomes from similarly configured DCF simulators. The DCF simulator in Chapter 5 is extended to simulate unsaturated stations described by ESN values. Statistics on the idle slot counts and the packet collision probability are collected and compared to model predictions.

6.2 Equivalent Saturated Nodes (ESN)

In this chapter, a station's degree of saturation is measured not by the probability of having something to transmit at a given slot time like [150] and [152], but by the proportion of all available transmission opportunities that is used by the unsaturated station to transmit.

Broadcast DCF decrements the backoff counter every slot time when the channel is idle, and resets to a value uniformly distributed within the contention window when the counter is zero. When the counter is reset, a saturated station transmits the frame in its buffer. The measure "Equivalent Saturated Nodes" (ESN) is defined as the probability that a station actually have something to transmit when the counter is reset. Since a completely saturated station will transmit every time it is able to, the probability it has something to send is 1 — such saturation is described as 1 ESN. For stations that, on average, only have something to send every two opportunities, they are 0.5 ESN. The concept of ESN is time-independent — it considers only transmission opportunities in the DCF backoff counter's perspective. It should be noted that actual implementations of the DCF do not continuously reset waiting for a frame — ESN assumes an "equivalent" modified station that continuously reset. This therefore assumes some average cycle length such that the probability can be calculated.

When using ESN to describe the contention level of a network, the degree of saturation for each station in the network is added. For example, a network containing 4 stations of 0.5 ESN has a channel contention level of 2 ESN.

6.3 Effects of station saturation

This section details the simulation study conducted to investigate the effects unsaturated stations have on channel observations.

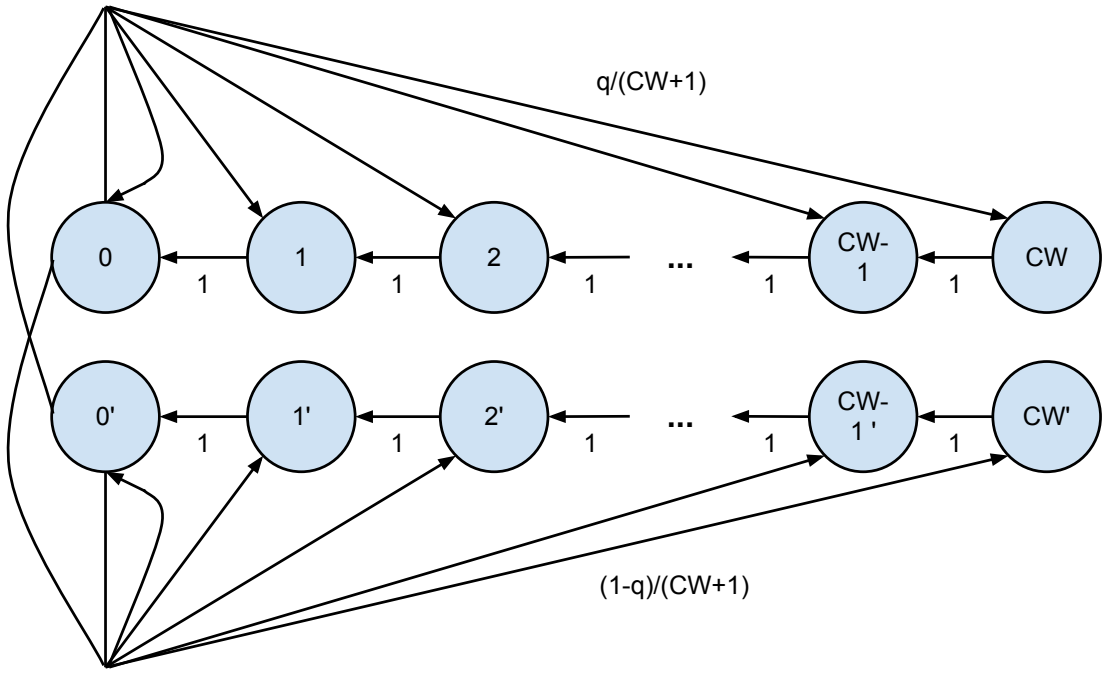


Figure 6.1: State machine description of the extended simulation model

6.3.1 Method

In this study, the simple DCF simulation model in Chapter 5 is extended such that each station can generate packets as specified by the ESN. In the updated DCF model, in addition to resetting the backoff counter to a uniformly distributed value within the contention window at the end of the backoff period, the station also chooses whether the station will put a packet onto the medium at the end of the next backoff period. The decision of whether to retransmit is as specified by the ESN value. Figure 6.1 is the state machine description of the updated simulation model. Unchanged from the original model, the updated model continues to use MRG32k3a as the pseudo-random number generator.

The simulation study conducted aims to identify:

- whether scenarios with the same total ESN behave similarly (*e.g.* 2 stations of 1 ESN vs 4 stations of 0.5 ESN); and
- whether and how accurately the original analytical model can predict MAC idle slot counts and collision statistics.

Table 6.1: Simulation parameters

CW size	31
Target Total ESN	$\{1, 2, 5, 10, 15, 20, 30, 40, 50, 100\}$
Station ESN	$\{1, .99, .95, .9, .75, .5, .3333, .25, .2, .125, .11111, .02\}$

To achieve these aims, the parameters specified in Table 6.1 are used for these simulations. The actual number of stations simulated and the actual total ESN are calculated from the parameters on execution of each simulation run. 500,000 idle periods from each simulation are collected and the results of the simulations are compared to the numerical solutions of the analytical model assuming saturated stations.

6.3.2 Results and discussion

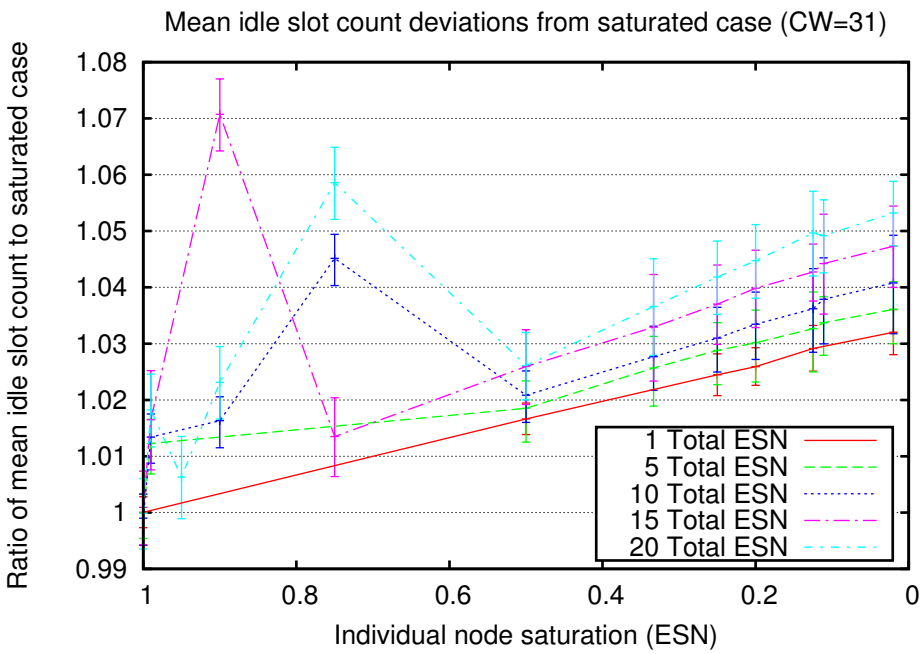
This section presents the results from the simulation, and discusses some of the issues immediately relevant to the results being presented.

Validity of the "Equivalent Saturated Node" metric

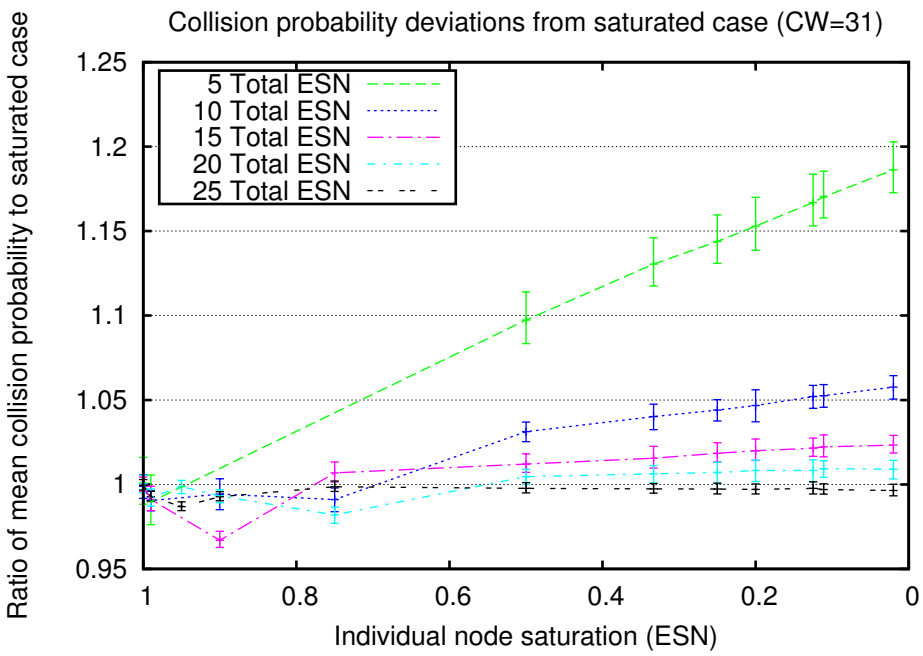
The mean idle slot counts and the collision probabilities from each simulation conducted are collected. Figure 6.2 plots the relationship between the changes in the observed idle slot count and the collision probability as a result of varying the station saturation.

As evident in these figures, station saturation does have a significant effect on both the idle slot counts and the collision probabilities. For low enough saturation, (ESN ≤ 0.5), the effect appears to be fairly consistent. A low station saturation increases the mean idle count and also increases collision probability. At higher station saturation, the effects seems to be relatively irregular.

The increase in the mean idle slots is intuitively obvious. As station saturation decreases, the probability that some (or many) stations have nothing to send increases, which is observed in the channel as idle slots.



(a) Mean idle slots deviations



(b) Collision probability deviations

Figure 6.2: Changes to mean idle slots and collision probability due to station saturation, $aCW_{min} = 31$. Plots show the ratio of the unsaturated (a) mean idle slot counts and (b) collision probability to their respective saturated values from simulation. Error bars shows two standard deviations from the sample means.

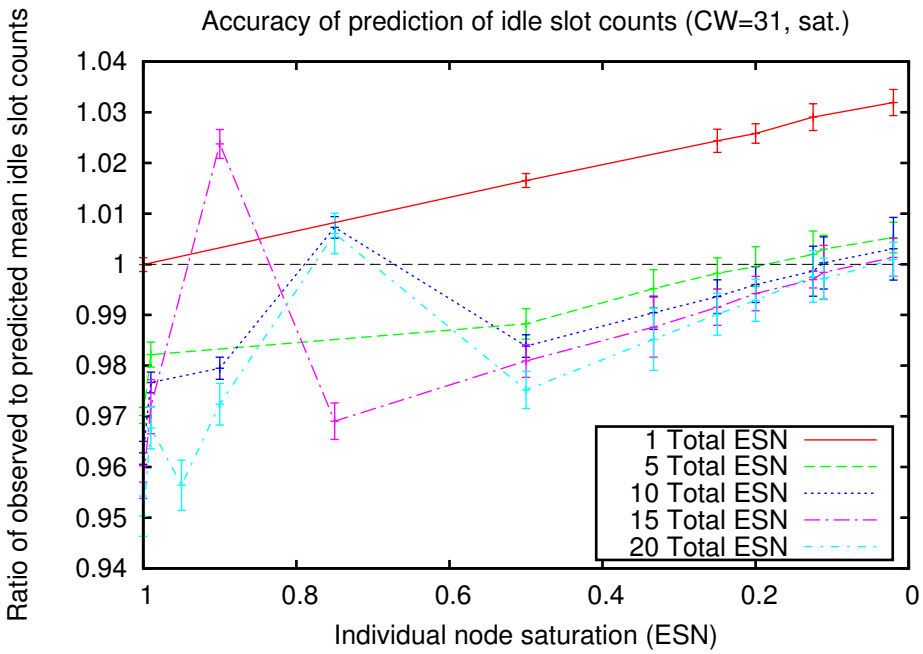
The increase in the collision probability can be explained by the increased number of stations on the network. Collision probability follows a binomial distribution. As the station count increases, both the exponent of the binomial distribution and the number of terms in the expression increase. Whereas as the saturation decreases, the base of the binomial distribution decreases. The net result is an increase in the probability of two (or more) stations contending for the same slot.

The irregular observations observed at higher station saturation, may be partially attributed to differences between the specified and the actual simulated network saturation. The differences observed are due to the fact that it may be impossible to generate the specified network saturation using the specified station saturation. (*e.g.* It is impossible to form a 1 ESN network using 0.75 ESN stations.) In these cases, the number of stations is rounded down, resulting in total network saturation below the value specified.

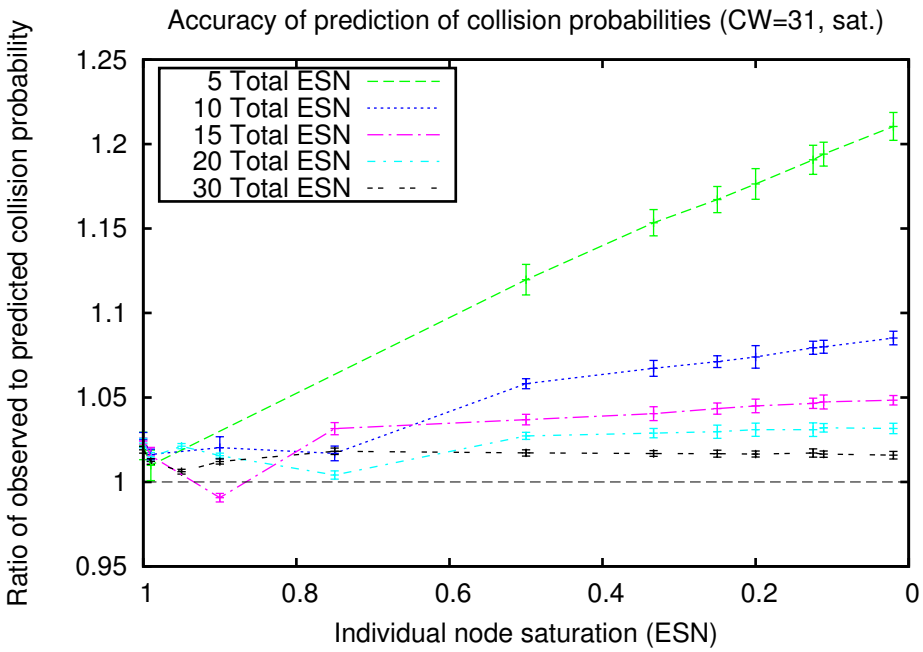
One can see that individual station's saturation does have an effect on both the mean idle slot counts and the collision probabilities. The magnitude of the difference on mean idle slot counts are relative small $\pm 8\%$, while the magnitude of the difference is up to $\pm 20\%$ at 5 ESN network saturation. Therefore, one can draw the conclusion that while the total network ESN does not fully describe the effects of the level of network contention, it can still provide a rough measure indicative of the underlying channel statistics.

Accuracy of theoretical models

Plotting the data obtained against the predictions of the saturated station model, the accuracy of the model in networks containing unsaturated stations can be gauged. In Figure 6.3, the ratio of the observed values to the predictions is plotted against individual station saturation. From the plot, one can see that, at station saturation (station ESN = 1), the model overestimates (ratio < 1) the mean idle slots counts (consistent with observations from Chapter 5) and under-



(a) Mean idle slots deviations



(b) Collision probability deviations

Figure 6.3: Accuracy of predictions made by the model assuming saturated stations, compared to simulation with various individual station saturation, $aCW_{min} = 31$. Plots show the ratio of the observed (a) mean idle slot counts and (b) collision probability to their respective predictions based on the original model assuming saturated stations. Error bars show two standard deviations from sample means.

estimates (ratio > 1) the collision probabilities. As station saturation decreases, the underestimation of collision probability becomes worse. On the other hand, the observed mean idle slot counts increase as station saturation decreases such that observed values approach, and in some cases exceed the predicted values. Note that the predicted values are not bounding the observed idle slot counts.

Based on the observed values, one can see that the theoretical predictions lie within $\pm 5\%$ for the mean idle slot counts (worst near saturation). For low enough station saturation, the effect of station saturation forms a linear relationship. Similar to the previous section, the unusual behaviour at high station saturation can be at least partially attributed to the simulated channel contention being different to the specified (thus plotted) value.

In terms of collision probability, the error is observed to grow linearly as the individual station saturation decreases (slight variations at high station saturation), and is worse for lower total network saturation.

6.4 Impact on channel estimation

Since individual station saturation has some effects on both the mean idle slot count and the collision probability of the system, the impacts of these differences on the channel contention measurement technique in Chapter 5 need to be investigated. In order to study these effects, similar to the previous chapter, a simulation study using two Bayesian inference-based channel load estimators as described in Chapter 5 is conducted.

6.4.1 Method

In this simulation experiment, the estimators are configured using the relevant probabilities from the DCF model assuming saturated stations. One of the estimators implements the channel busy status observation technique presented by Bianchi and Tinnirello [136], while the other implements the interframe idle slot

count technique in Chapter 5. While keeping the network's total ESN constant, the individual ESN of stations are adjusted and extra stations are added as required. It is noted that the network's total ESN cannot be maintained for all values of station saturation — *e.g.* it is impossible to form a network of 1 ESN when all stations are 0.99 ESN saturated. In these situations, the next lower integer number of stations are used as long as the contention of the network formed is within ±5% of the specified value.

In this experiment, statistics on the time to reach steady state as well as the steady state values are collected. The estimator is considered to have reached steady state if its belief vector has not changed by more than 10^{-9} cumulatively for at least 1000 updates. The time to reach steady state is therefore one after the last time step that caused the belief vector to change by more than the threshold. Furthermore, once the steady state had been reached by both estimators, the experiment is terminated.

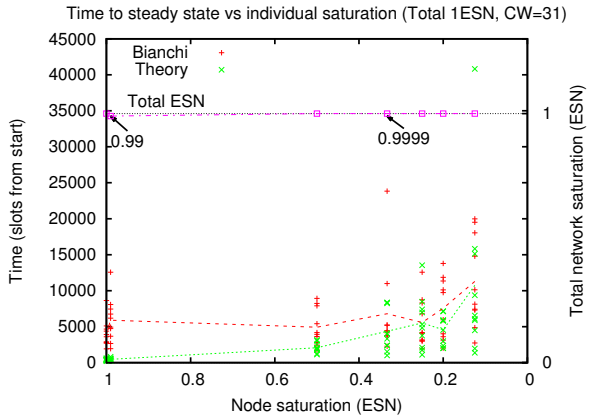
Ten executions of the simulation using different random seeds were run. The pseudo-random number generator used in these simulations is also MRG32k3a.

6.4.2 Results

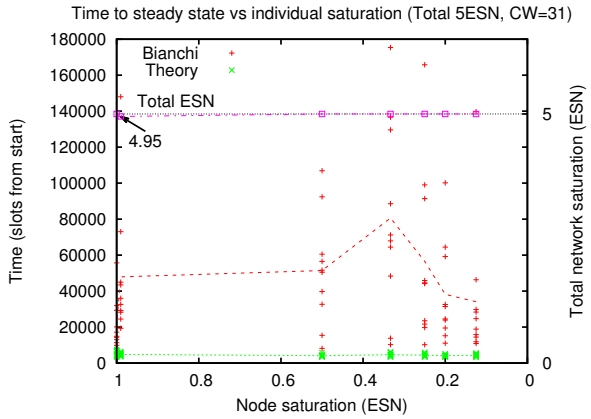
Time to converge

Based on the trace of estimator belief vector, the time to converge to steady state for each estimator and each scenario can be analysed. Figure 6.4 plots the time to converge against individual station saturation.

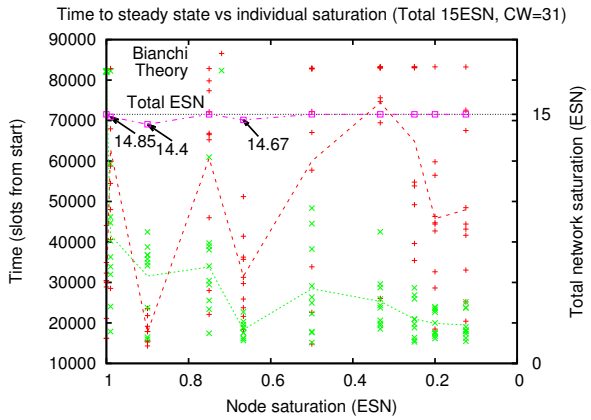
Overall, the figure suggests that the technique presented in Chapter 5 tends to converge to steady state faster than by observing channel busy status [136] for most cases. The technique of observing interframe idle slots consistently perform better across the various individual station saturation levels, except of one case (0.9 ESN in Figure 6.4c). This data point also corresponds to the situation where the actual simulated contention level is lower than specified. Most dips in the performance of the idle slot technique correspond to cases where large variations



(a) 1 network total ESN



(b) 5 network total ESN



(c) 15 network total ESN

Figure 6.4: Effect of station saturation on the time for Bayesian estimators to converge to steady state. Diagrams plot the time taken to converge for both Bayesian estimators configured for Bianchi and Tinnirello [136] and the DCF model assuming saturated stations (Chapter 5) for each execution of the scenarios. Red and green lines represents the respective means for the two approaches. Purple lines (right hand axis) mark the total network ESN, with the scenarios that are not exactly as specified in the title marked. Steady state is defined as estimator belief vector not changed by more than 10^{-9} cumulatively for 1000 updates.

existed between the simulated and specified values. This behaviour is consistent with the previous findings on estimating non-referenced values.

Accuracy of steady state estimates

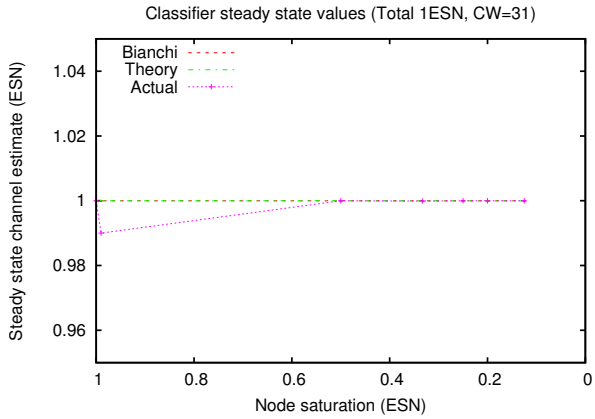
The accuracy of the techniques is assessed by plotting the final steady state estimate from estimators against the individual station saturation in Figure 6.5. The final steady state estimates are taken as the weighted mean of the belief vectors. The figure shows the min, median, max and the interquartile ranges of the estimates from the 10 runs executed for each scenario.

These results suggest that at low total network saturation (1 ESN, Figure 6.5a), the effect of individual station saturation is not significant between the two techniques. This may be because at low channel contention, the observation probabilities for both idle slot counts and packet collisions are very distinctive such that the small variations caused by unsaturated stations are not significant.

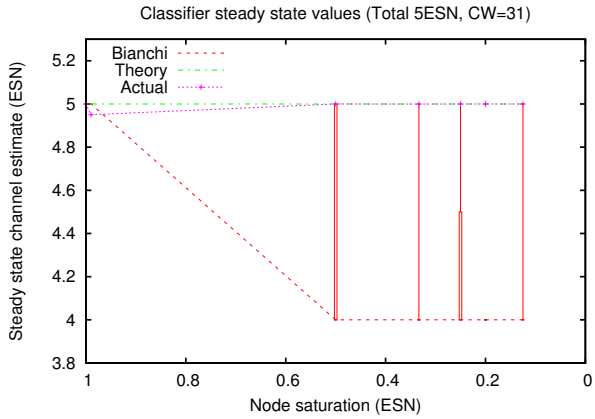
As the total network contention increases, the effect of unsaturated stations increases. Decreases in station saturation lower the contention estimate for both techniques, with the effect more prominent for method observing channel busy status. Figure 6.5c shows an overestimation for the estimator observing idle slot counts. The overestimation is not actually caused by the decreased station saturation — this effect is also observed with saturated stations. The results show that observing idle slots produces more accurate results than observing channel busy status in most scenarios tested. These plots also suggest that the method of observing idle slots also cause the estimator to be more certain about its estimates than Bianchi and Tinnirello's technique.

6.5 Accounting for saturation in the DCF model

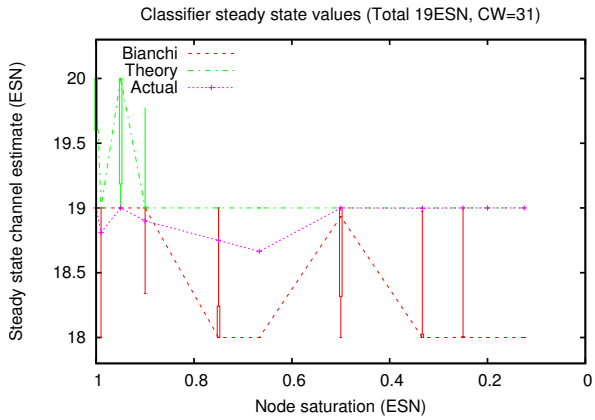
Notwithstanding the ability for the channel estimation techniques to operate relatively accurately in presence of unsaturated stations, these technique may be able to be more accurate if the DCF model underlying the technique can be improved,



(a) 1 network total ESN



(b) 5 network total ESN



(c) 15 network total ESN

Figure 6.5: Effect of station saturation on the final steady state estimates from the Bayesian estimators. These box-and-whisker diagrams plot the spread of the weighted mean belief vectors, plotting the min, max, median and interquartile range of the observed final estimate with estimators configured for Bianchi and Tinnirello's [136] and the DCF model assuming saturated stations (Chapter 5). Red and green lines represents the respective median estimates from the two approaches. Purple lines are the actual total network ESN (some are not exactly as specified in the title). Steady state is defined as estimator belief vector not changed by more than 10^{-9} cumulatively for 1000 updates.

accounting for unsaturated stations.

Unsaturated stations have been modelled in a variety of ways previously. One such method is to incorporate transmission probability by introducing extra "waiting" states into existing Markov models. Daneshgaran *et al.* [150] added an extra idle state that optionally precedes the actual backoff states to represent the time when the station has nothing to send. The station moves to this idle state only on completion of the previous backoff with probability $1 - q$, remains in this idle state with probability $1 - q$ and moves out of the state into one of the backoff counts with probability q/CW. Malone *et al.* [152] instead adds a parallel backoff states above the Bianchi model allowing stations to continue to count down backoff slots even when it has nothing to send. The station has a probability q to transit back to the normal initial backoff chain at the next lower slot. (*i.e.* Packets arrive with probability q at each slot time.) Both of these methods model saturation as a probability of packet arrival.

The concept of "Equivalent Saturated Nodes" used in this chapter does not model packet arrival, and can be easily added to the Markov model for saturated stations presented in Chapter 5. The extension involves adding an identical parallel "not sending" chain such that at the completion of packet transmission (the original state 0), the station moves to a state in the original chain with probability q, and to the new "not sending" chain with probability $1 - q$ (Figure 6.6). Macroscopically, the station will act as though it is saturated, but transmitting only q out of all the opportunity for it to transmit. Since the aim is to determine the number of slots between transmissions, which of these two chains a station is in at a given time is unimportant. Therefore the model can be simplified by merging the equivalent stations between each chain, and distinguish only between the "real" state 0 and the "not sending" state 0 (Figure 6.7).

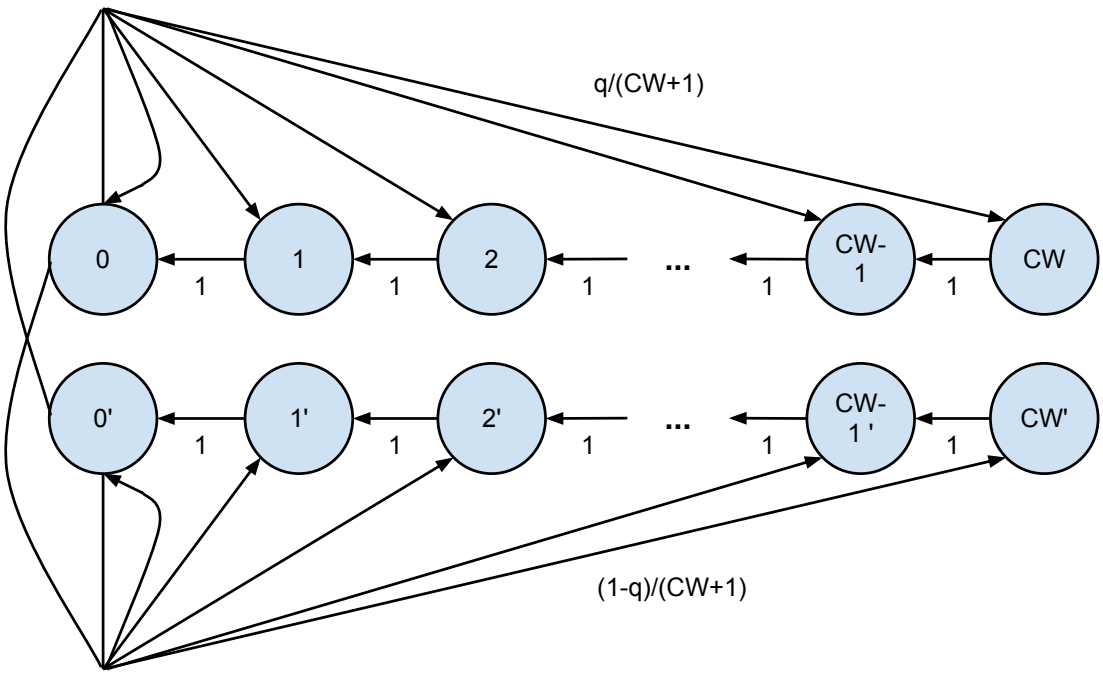


Figure 6.6: Markov model accounting for ESN value of the station. The bottom row of states are corresponding to backoff slots that, once finished count down, will not result in an actual transmission. The variable q is the probability that the next cycle will result in a transmission.

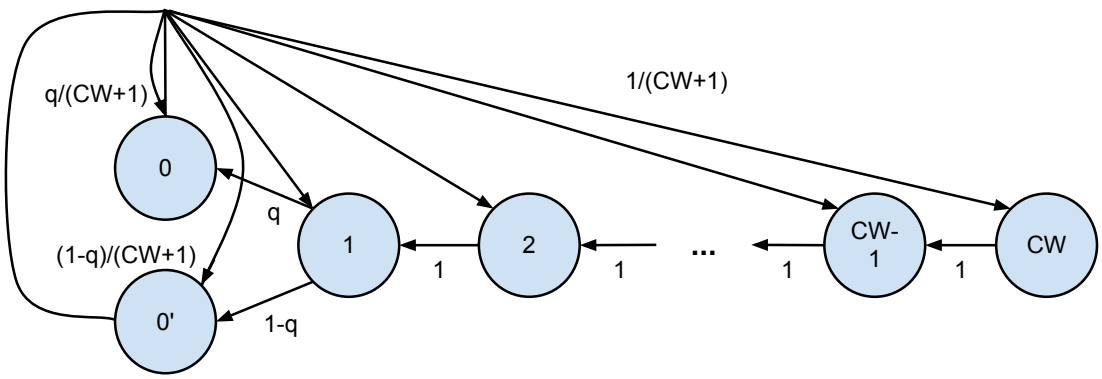


Figure 6.7: Simplified model merging equivalent states. Here, the non-zero backoff slots from both the sending and the non-sending chains are merged because, for the purpose of our evaluation, they are equivalent.

Steady State Expression

Using the same process as Section 5.2, the steady state solution S_k to the simplified model (Figure 6.7) is trivial to derive.

$$S_k = \begin{cases} (CW+1-k)\,\frac{P_0}{CW+1} & k \in [1, CW] \\ q\,(CW+1-k)\,\frac{P_0}{CW+1} & k = 0 \\ (1-q)\,(CW+1-k)\,\frac{P_0}{CW+1} & k \text{ is } X \\ & X \text{ is the non-sending state } 0 \end{cases} \tag{6.1}$$

Unlike Section 5.2, the length of time a station waits before transmission cannot be determined simply from the steady state solution. For further analysis, the probability that a station is k slots away from transmitting needs to be determined. *i.e.* merging the probabilities of all state transition that is k slots away from state 0, resolving any loops caused by one or visits through state X. Such linearisation will yield an infinite sequence of (mostly) descending probabilities.

In order to generate such mapping, one can form a transition matrix $\mathbb{T}$ based on the simplified Markov model, and removing all the outbound arcs from state 0. The mapping can be computed numerically by computing $\mathbb{S}\mathbb{T}^k$, where $\mathbb{S}$ is a vector containing the steady state solution to the Markov model, and inspecting the element corresponding to state 0.

$$P_k = (\mathbb{S}\mathbb{T}^k)_{\text{state0}} \tag{6.2}$$

A finite approximation of the function P_k can be obtained by truncating the series after $\sum_0^k P_k > 1 - \epsilon$ for some small threshold value ϵ. The series should than be scaled such that the total probability equals 1.

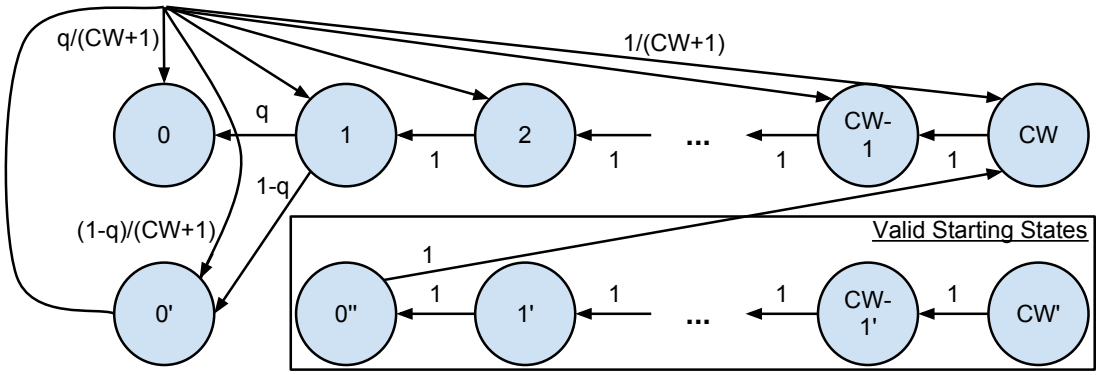


Figure 6.8: State machine for the transition matrix of the group of stations with nothing to send currently.

6.5.1 Relationship between unsaturated station count and interframe period

A similar expression to Section 5.3.2 can be derived to improve the fidelity of the model, but it is unfortunately more complex. In Section 5.3.2, all stations who had previously transmitted participate in channel contention, and choose a slot according to the uniform distribution. In the unsaturated situation, this is not necessarily the case. Therefore, a further partition of the last transmitted stations is required — those who has something to transmit (hence chooses slots uniformly distributed in $[0, aCWmin]$, and those who don't. In order to simplify the final expression, the current definition of U_k is extended by expanding the domain of the function:

$$U_k = \begin{cases} \frac{1}{CW+1} & k \in [0, CW] \\ 0 & \text{otherwise} \end{cases} \tag{6.3}$$

Note that U_k is independent of q (the probability of having something to transmit). This is because U_k is used only the stations that transmitted and have something to transmit — the probability of whether it has something to send is accounted for outside this expression.

In addition to redefining U_k, a new distribution D_k is needed for the stations who don't currently have frames to send. Similar to P_k, D_k is also evaluated

numerically from a transition matrix. Since D_k is used by stations that have nothing to send, a parallel "definitely nothing to send" chain is needed (Figure 6.8). After their last transmission, these stations follow the normal DCF behaviour, selecting a backoff that is uniformly distributed along the "definitely nothing to send" chain. This chain finishes at state X, after which the station returns to the standard DCF model. Numeric solution of D_k will yield another In finitely long series, and can be truncated and scaled in the same way as P_k to yield a finite approximation.

Finally, expressions similar to Section 5.3.2 can be derived:

$$
\begin{aligned}
{}^{0}T_0 = & \sum_{i=1}^{N} {}^{i}T \binom{N-i}{0} (1-P_0)^{N-i} \\
& \sum_{d=0}^{i} \binom{i}{d} q^{i-d} (1-q)^d (1-U_0)^{i-d} (1-D_0)^d \qquad (6.4) \\
{}^{n}T_{k|{}^{0}T_{k-1}} = & \sum_{i=1}^{N} {}^{i}T \sum_{m=\max(0,n-N+i)}^{\min(i,n)} \binom{N-i}{n-m} \tilde{P}_k^{\,n-m} \left(1-\tilde{P}_k\right)^{N-i-n+m} \\
& \sum_{d=0}^{i} \binom{i}{d} q^{i-d} (1-q)^d \\
& \sum_{h=\max(0,m-d)}^{\min(i-d,m)} \binom{i-d}{h} \tilde{U_k}^{\,h} \left(1-\tilde{U_k}\right)^{i-d-h} \\
& \binom{d}{m-h} \tilde{D_k}^{\,m-h} \left(1-\tilde{D_k}\right)^{d-m+h} \qquad (6.5) \\
\tilde{X_k} = & \frac{X_k}{1-\sum_{m=0}^{k-1} X_m} \qquad X \in \{D, P, U\} \qquad (6.6)
\end{aligned}
$$

6.6 Evaluation of the extended model

To assess the validity of the extended model, a simulation study comparing the model predictions to simulation outcomes is conducted.

Table 6.2: Simulation parameters

CW size	31
Target Total ESN	$\{1, 2, 5, 10, 15, 20, 30, 40\}$
Station ESN	$\{1, .99, .95, .9, .75, .6667, .5, .3333, .25, .2, .125\}$

6.6.1 Method

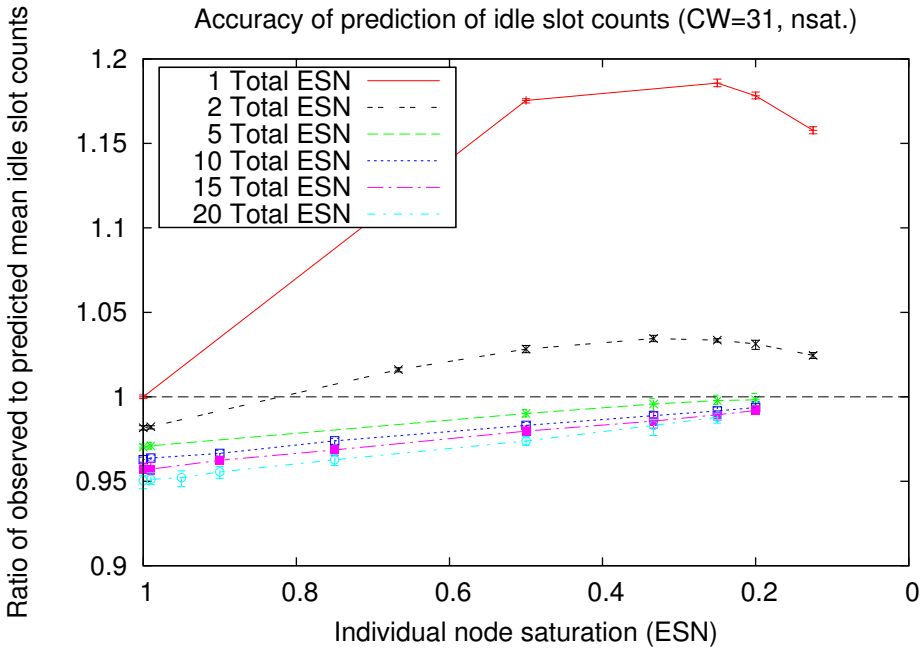
The simulation study uses the extended DCF simulation model in Section 6.3 as the "ground truth" to compare the model predictions to. In these evaluations, the scenarios in Table 6.2 are configured for both the simulation and the theoretic model. Actual number of stations assessed and the actual total network ESN are calculated from the parameters on execution of the simulation. It is noted that the parameters in Table 6.2 may specify certain total network ESN values that is impossible to achieve for the given station saturation — *e.g.* total 1 ESN for stations that are at 0.6667 ESN saturation. In these cases, the next lower integer number of stations are used as long as the total network contention is within ±5% of the specified value, resulting in a total network contention that may be slightly below the values specified. Numeric solutions to the theoretic model are obtained using the algorithm presented in Chapter 5.

500,000 idle periods were collected from each simulation. The results of the simulation are then compared to numerical solutions of the analytical model.

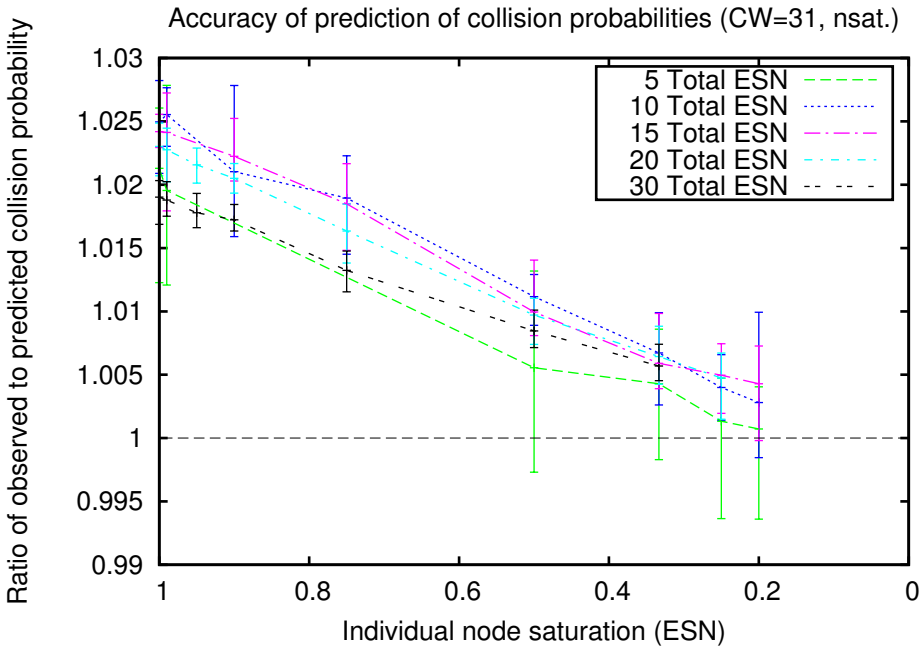
6.6.2 Results and discussion

Figure 6.9 shows the accuracy of the extended model. At 1 total ESN across the network, the results show a maximum error of 18.8% at 0.33 individual station ESN for mean idle slot counts. For networks with higher total ESN, the results recorded lie well within ±5%.

In terms of collision probabilities, the predicted results lie within ±3%, with the model underestimating the collision probabilities. The error decreases approximately linearly as individual station saturation decreases. Furthermore, it is observed that the prediction error increases as channel saturation increases un-



(a) Mean idle slots deviations



(b) Collision probability deviations

Figure 6.9: Accuracy of predictions made by the model accounting for station saturation, compared with simulation with various individual station saturation, $aCW_{min} = 31$. Plots show the ratio of the observed (a) mean idle slot counts and (b) collision probability to their respective predictions based on the extended model.

til some maximum before reducing. Due to the excessively long computation time required to solve the model, no data have been collected for channel contention above 40 ESN or station saturation below 0.20 ESN (except for 1 and 2 ESN channel contention where data is not available for saturation below 0.125 ESN).

The lack of data for station saturation under 0.2 ESN or channel contention over 40 ESN demonstrates the poor scalability of the model solving process. The algorithm used to compute model predictions is extremely computationally intensive. Compared to the original model, the extended model's complexity has grown by another magnitude in n due to the extra nested summation in the equation. Numerically evaluating for 50×0.02 ESN stations (1 total ESN) required more than 2 hours, and was terminated before completion. In addition, the error in P_k after the solver exited is at times more than 1%. (The solver is unable to iteratively improve on the error.) For this model to be practically useful, better numeric evaluation techniques and/or simplification of the equations are required.

Comparing the extended model with the original model in Figure 6.3b, one can see that the extended model gives a better prediction of collision probabilities with the extended model underestimating by less than 3% compared to 20% error at 1 ESN in the original model. All data points collected show that this model underestimates on average, but is not conclusive due to the lack of data points at the extremity. In addition, apart from the 1 ESN case, the model predictions are within ±8% (±0.1 slots) from the observed value. This performance, in terms of the magnitude of error, is similar to those observed in the original model, but the original model underestimates whereas the extended model tends to overestimate at high channel contention.

6.7 Conclusion

In this chapter, the effects of homogeneous unsaturated stations on the network is studied. It is found that the measure "Equivalent Saturated Nodes" (ESN)

can be used to describe both the degree of saturation of a given station and the total amount of contention in a network, but is insufficient to completely characterise the network load. Unsaturated stations on a network is shown to have observable non-linear effects on both packet collision ratio and observed idle slot counts. Decreases in individual station saturation result in both higher mean idle slot count and higher collision probability compared to the saturated case. For networks or stations that are saturated enough, the differences are small. This suggests that while the ESN measure can be used to describe very saturated situations, it does not fully capture the effect of station saturation on channel observation.

The validity of the DCF model assuming saturated stations, as described in Chapter 5, was tested against various station saturation. The saturated station model is able to predict the idle slot counts fairly accurately ($\pm 5\%$), but greatly underestimates the collision probability at low station saturations.

It is shown that the technique of observing interframe idle periods to estimate channel contention can be used on networks containing homogeneous unsaturated stations when there are no hidden terminals. Simulation of the operation of both Bianchi and Tinnirello's technique [136] and the idle slot technique showed that both are capable of estimating the total network load in terms of "Equivalent Saturated Nodes" (ESN), but unsaturated stations cause the estimators to take longer time to converge to the steady state. Additionally unsaturated stations also cause an underestimation of ESN compared to the saturated case.

An extension to the DCF model incorporating unsaturated stations was also presented, and was shown to be not scalable in its current form. The numerical evaluation technique used is not sufficiently accurate, and it still takes too long to compute due to the extra nested summation in the extended model. In terms of accuracy, the extended model is able to predict the packet collision ratio quite accurately ($\pm 3\%$), but differs greatly from the observed values when the total network saturation is low.

In the next chapter, the accuracy of using the ESN metric to predict packet reception is assessed. Given the accuracy of this metric, one can produce a geocast system that is not only efficient in channel use, but can also adapt to channel contention automatically by combining the work presented thus far.

Overview

This chapter combines the techniques presented in previous chapters into a load-reactive geocasting system.

Contributions

- **I have validated the usefulness of the ESN metric in predicting packet non-reception.** Statistical analysis on computer simulation results showed that a simple threshold test on observed ESN value has a very high Negative Predictive Value. This means that ESN can very accurately predict packet non-reception.

- **I have used statistical techniques to provide further evidence on the efficiency of the interference-aware geocast algorithm.** Statistics on the ESN-based threshold test shows that the test has a higher Positive Predictive Value on the interference-aware geocast algorithm than a greedy distance-based technique. This provides further evidence that the geocast algorithm is more efficient in using the channel to improve packet reception.

- **I have designed and evaluated a geocast algorithm that changes its behaviour in reaction to channel contention.** This algorithm uses outputs from the passive idle slot-based channel estimator to determine whether rebroadcasts should be increased or suppressed, and adjusts the retransmission parameter of the interference-aware geocasting algorithm automatically. This allows the algorithm to adapt to channel conditions without the need for manual intervention.

Chapter 7

Load-Reactive Geocasting

7.1 Introduction

Having introduced the technique to obtain passive channel contention estimates in Chapter 5, this chapter now applies this technique to the interference-aware geocast algorithm presented in Chapter 4 in order to allow the algorithm to automatically adapt to channel load.

In this chapter, the use of the "Equivalent Saturated Node" (ESN) metric to predict packet non-reception in situations with both hidden terminals and unsaturated stations is first validated. Second, this ESN metric is used in the design of a load-reactive geocast system, coupling the channel contention estimator to the interference-aware geocast algorithm. Through an ns-3 based computer simulation, the geocast algorithm is shown to be effective in controlling the channel load introduced by retransmissions, hence improving packet reception over a wide range of vehicle densities without manual intervention. Emergent behaviours observed from this two-part system are also described and discussed.

7.2 Determining optimal channel conditions

Results from the geocast simulations in Chapter 4 suggest that packet reception ratio (PRR) is highly dependent on channel load. It is also shown that the PRR

can be improved by cleverly selecting rebroadcasters and encouraging retransmissions up to some threshold channel saturation level. Furthermore, as long as the channel is not too busy, retransmissions add redundancy, allowing stations that originally failed to receive the packet to receive a forwarded copy. By knowing the current channel contention, reception can be improved by either increasing or decreasing retransmissions as appropriate.

The selection of an appropriate channel contention metric is therefore important for a load-reactive geocast algorithm to determine the amount of contention, hence the probable PRR, of a channel. Chapter 4 used the channel busy time (CBT) as the metric to the describe channel load, and showed that an extremely high CBT is associated with low PRR. Unfortunately, the mapping between CBT and PRR is not consistent, and is therefore an unreliable measure to base a load-reactive algorithm on. An alternative passive channel load estimation technique was presented and evaluated in Chapters 5 and 6, and it was demonstrated that, in ideal channels, this technique can predict PRR very well.

In this section, the relationship between the channel load and the PRR is further investigated, with the aim of finding a good measure of channel load for the load-reactive geocasting algorithm. By implementing the passive channel load estimation technique into the ns-3 simulation in Chapter 5 and collecting the load measurements in the simulation, it was shown that the "Equivalent Saturated Node" (ESN) metric is a useful load measure for our purpose. Furthermore, an optimal ESN value for predicting packet non-reception is also determined.

7.2.1 Method

The relationship between channel load and PRR is determined empirically through computer simulations using the ns-3.9 simulator. Since the investigations in Chapter 4 show an abrupt drop in mean PRR as vehicle density is increased, a much finer increment in vehicle densities is used in these simulations in an attempt to identify the nature of the transition from good to bad PRR.

Table 7.1: Simulation parameters for selection of load measures

Vehicle densities	{10, 15, 20, 25, 30, 45, 50, 60} veh/km/lane
Proportion of HV	10%
Lanes per road	6
Road length	900 m
Intended coverage radius	200 m
Position of source	Centre of intersection
Antenna Gain (Tx and Rx)	2.512 dB
Rx Threshold	-95 dBm
CS Threshold	-99 dBm
Log-Distant Exponent (γ)	2.0
Log-Distant Ref Loss (at 1 m	-47.8588 dB (Friis loss, 5.9 GHz)
Nakagami parameter (m)	5.0
Receiver Noise	0 dB
Attenuation across HV	-20 dB
Building Shadow	-30 dB
Max transmission range	{50, 100, 150, 200}m
Transmission rate	6 Mbps
Transmission bandwidth	10 MHz
SIFS	32 μs
Slot Time	13 μs
Packet Rate	10 Hz
Packet lifetime	100 ms
Packet size	54 octets (incl. all headers)
Retransmission algorithms	{Greedy distance-based, interference-aware}
Retransmission parameter X	{0.5, 1, 2, 3, ..., 10}
Number of transmitters	All stations

In this experiment, a load estimator implementing the idle slot channel observation technique is attached to each station on the network. The probability distributions for idle slot counts for various numbers of saturated stations are precomputed and inserted into each estimator. Similar to the investigation in Chapter 6, idle slot counts from the MAC layer are fed to the load estimators when it senses the channel has switched from the IDLE state to any busy state.

It is important to note that the IEEE 802.11 implementation in ns-3.9 contains a known bug ("DCF Immediate Access" bug [186]), whose implications on MAC-layer measurements are detailed in Chapter 8. In order to account for this non-

standard behaviour in the simulator, the idle slot probability distributions are precomputed for a contention window size of 15 (one less than specified in the standard) and the DIFS is assumed to be 3 slot times (instead of 2).

Station layouts as well as most other simulation parameters are identical to those used in Chapter 4, except the vehicle densities investigated are more fine-grained. Detailed explanation of these parameters are contained in Chapter 4. Table 7.1 summaries the values of the parameters used.

In these simulations, the various channel contention measures were tested on both a greedy distance-based forwarding algorithm by Briesemeister *et al.* [84] and the interference-aware geocast algorithm in Chapter 4. The interference-aware geocast algorithm was simulated using a range of fixed values for the re-transmission parameter (X) in order to observe the estimator's behaviour over a range of contention levels.

Statistics collected include the channel busy time (CBT) as a proportion of the total simulation time, the packet reception ratio (PRR), the last estimated channel load at the conclusion of the simulation and the proportion of observed idle slot counts that are above one contention window. To minimise the impact of edge effects, only the packets sent from one tagged vehicle positioned at the centre of the simulated field is considered when computing the PRR. In addition, statistics are gathered only from stations within the target area of the tagged station for the same reason. The simulation field is large enough such that edge effects on the stations in the target area are minimal. The final state of the belief vector for each station is also recorded.

7.2.2 Results

This section presents results from this simulation study. Some results and their implications are also further discussed within this section.

Relationship between channel measures and vehicle density

This investigation compares three channel measures — the channel busy time (CBT), the estimated equivalent number of saturated stations ("Load") in ESN, and the proportion of inter-frame idle slot counts being above one contention window length (*"over CW"*). CBT is the proportion of time a station is either transmitting, receiving or has sensed the channel to be busy, to the total simulation time; the estimated load is obtained from the channel estimator implementing the idle slot count technique. Results presented are the final values at the conclusion of the simulation. These measures are collected at each station, and, unless otherwise stated, they are presented without further aggregation with other stations in the same scenario. Figure 7.1 shows the relationships between these measures and the configured vehicle density when the greedy distance-based algorithm is used.

From Figure 7.1, one can see that all three measures vary as a monotonic function of vehicle density on average, but their final values can have high variance for each vehicle density setting. The figures on the left plot the scenarios where all stations transmit with sufficient power such that, in the absence of shadowing, fast fading or packet collision, all stations within the target area can receive the packet ($S = 1$). These plots suggest that both the mean and the median of all the measures investigated change monotonically up to 25 vehicles per km per lane, after which they appear to flatten. This flattening suggests that channel saturation had been reached and the measures used are unable to precisely describe the actual load offered to the network.

When the transmission power of each station is lowered, the area experiencing interference introduced by each station reduces, thus the contention on the channel is lower. The plots on the right in Figure 7.1 show a corresponding reduction (or increase in the case of *"over CW"*) as a result of reducing the transmission range to one quarter of the required radius. It is noteworthy that Figure 7.1d shows significant outliers with the estimated ESN value when the station density

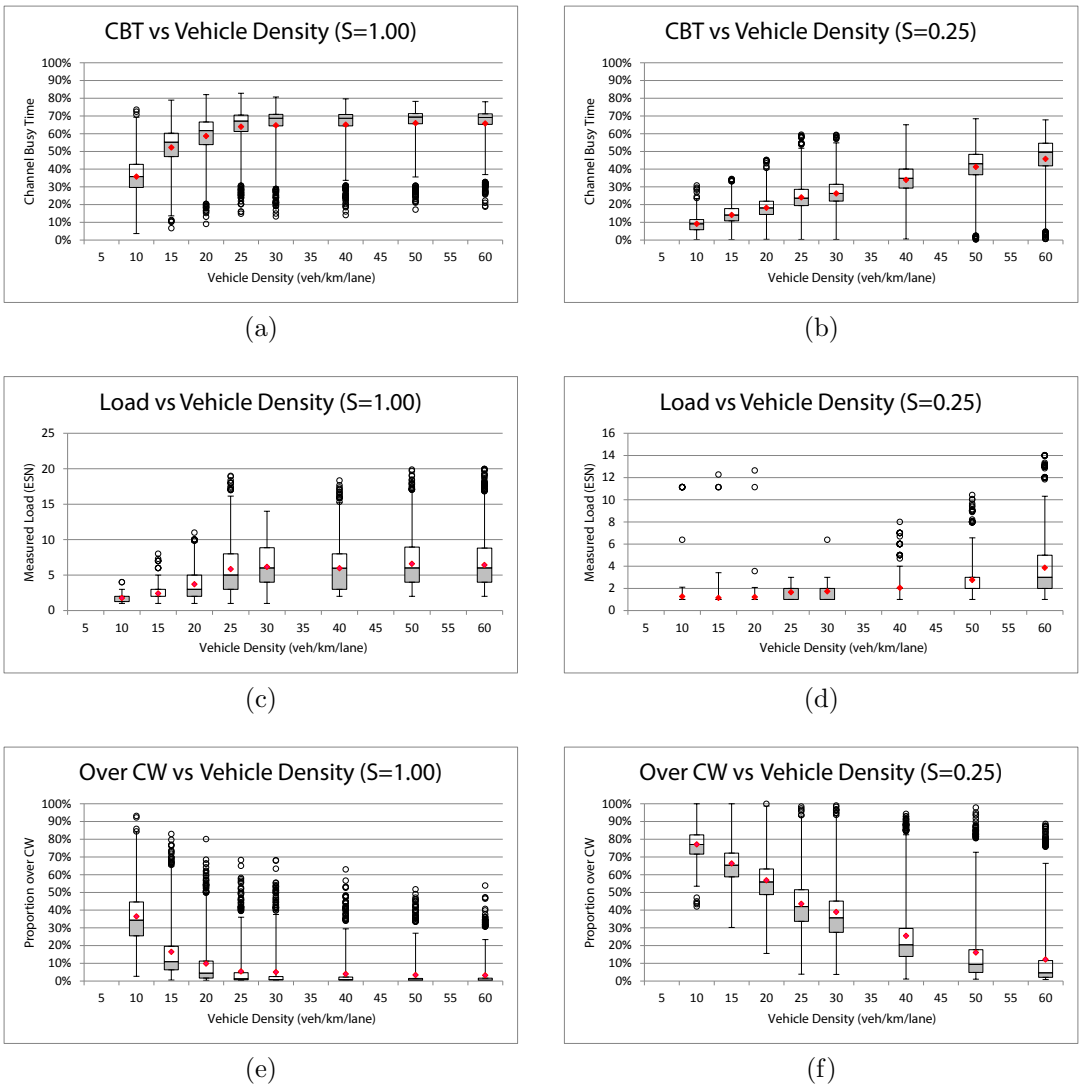


Figure 7.1: Results of centre-of-intersection scenario, all stations transmitting and use greedy distance-based forwarding algorithm. The body of the box-and-whiskers plot shows the inter-quartile range with the means marked by purple diamonds. Whiskers extend to the furthest observations within ±3 standard deviations from the mean. Outliers close to the end of the whiskers might have been omitted due to software limitations. S is the ratio of the maximum transmission range to the required coverage radius.

is very low.

Relationship amongst the channel measures

Interestingly, if the load measures observed at each station is plotted against the other measures from the same station, insights on the measures' relationships with each other can be gleamed. Figure 7.2 contains scatter plots between each pair of the measures at each station. Both CBT and *"over CW"* show a step-like

relationship against ESN load — flat at high ESN, and almost vertical at low ESN. It is noteworthy that the outliers at low load from Figure 7.1 can also be seen in Figures 7.2b (spurious high ESN at 0% CBT) and 7.2d (spurious high ESN near 100% *"over CW"*).

CBT and *"over CW"* show an almost linear relationship until the load becomes very high, showing a slight flattening when the CBT is greater than approximately 60%. It should be noted that in the plots presented, data points are overlaid on top of each other by series, making certain regions in Figure 7.2f appear concave. The visible area for each vehicle densities (except for the case where the density is 10 vehicles/km/lane) represents only the lower and leftmost bounds of the region — the top edge of the region is the same as the upper bound for the blue region (10 vehicles/km/lane) and are obscured by these data points. When the vehicle density increases in these plots, the regions are simply widened and lengthened with little changes in concavity.

Predictive value of the measures

By plotting the packet reception ratio (PRR) against the channel measures, one can ascertain whether these measures are useful for predicting packet reception. Figure 7.3 shows the relationship between these measures and the corresponding PRR at each station.

The top four plots (especially Figures 7.3b and 7.3d) shows both high and low PRR across the range of observed metric values, suggesting that PRR is not highly correlated to either CBT or *"over CW"* notwithstanding that associations between mean CBT and mean PRR were observed in Chapter 4. Figures 7.3e and 7.3f show an almost step-like shape, with high estimated ESN appears predictive of non-reception of the packet. It should however be noted that low ESN is not indicative of good PRR. The usefulness of the ESN for PRR prediction is further validated in Section 7.2.3.

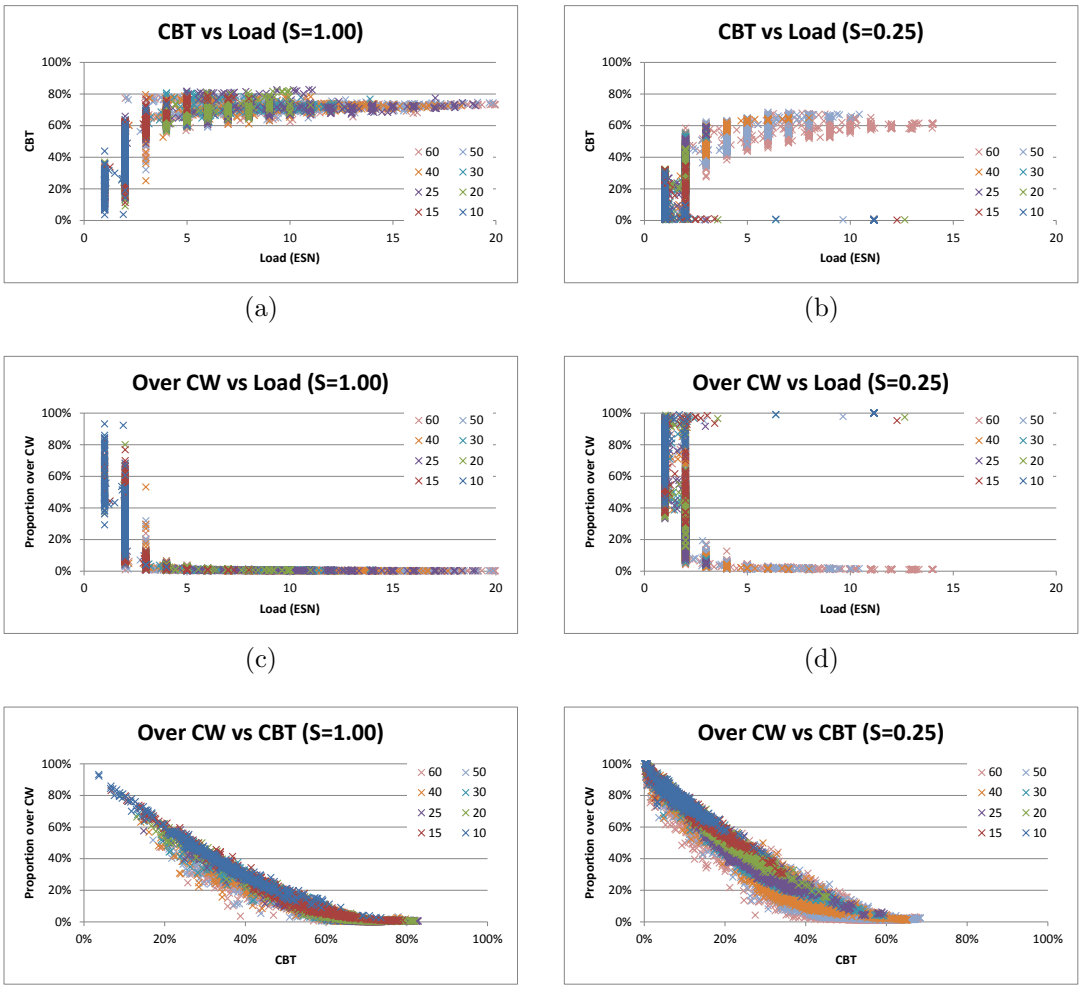


Figure 7.2: Scatter plot amongst the three measures for each station in the scenario. Centre-of-intersection scenario, all stations transmitting and use greedy distance-based forwarding algorithm. S is the ratio of the maximum transmission range to required coverage radius. It should be noted that there are significantly more data points for the higher vehicle density scenarios because there are more stations. The colour of the points depicts the vehicle density configuration the data point is sourced — the units are in vehicles per km per lane.

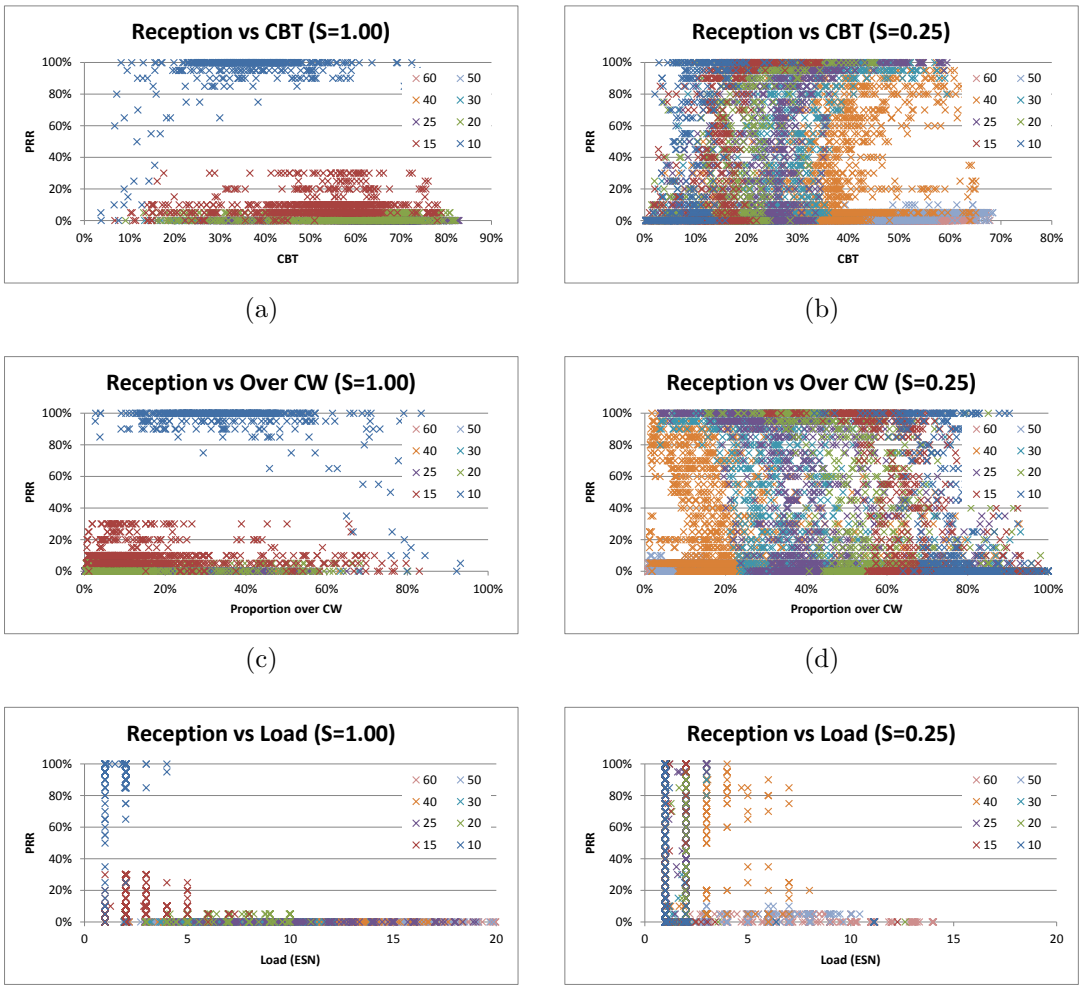
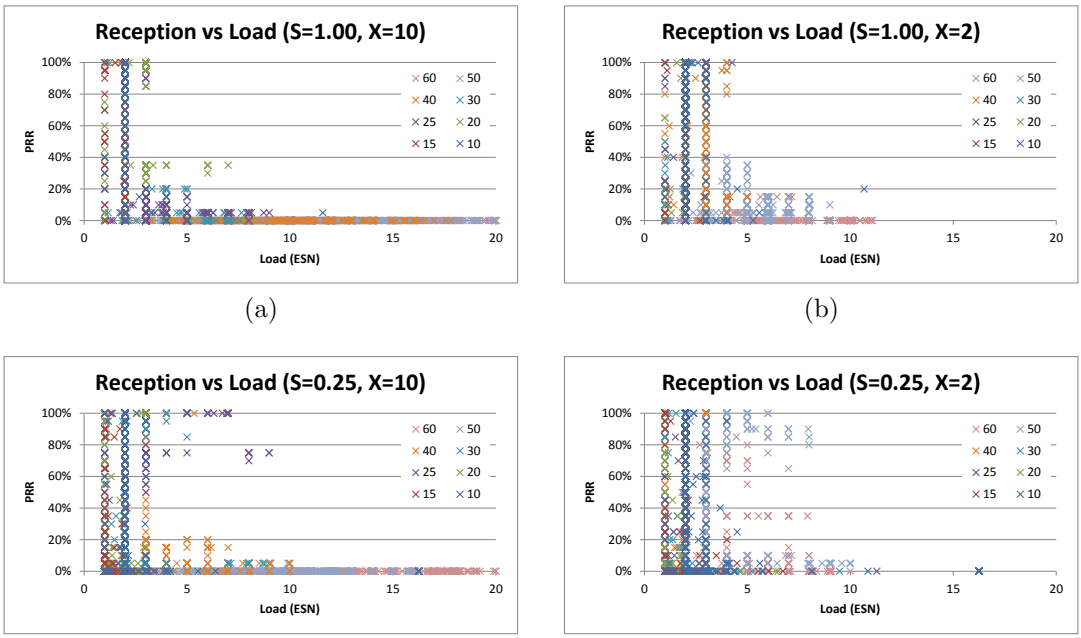


Figure 7.3: Scatter plots illustrating the potential predictive power of the various channel measures. Results from centre-of-intersection scenarios, all stations transmitting and use greedy distance-based forwarding algorithm. S is the ratio of the maximum transmission range to required coverage radius. The colour of the points represents the vehicle density configuration the data point is sourced — the units are in vehicles per km per lane.

Differences with the interference-aware geocast algorithm

As expected, the PRR-vs-Load scatter plots (Figure 7.4) does not show significant differences between the greedy distance-based algorithm and interference-aware geocast. This confirms the utility of the ESN-based load value as a measure of channel contention in predicting packet non-reception for both the greedy distance-based and the interference-aware geocast algorithms.



(a) (b)

(c) (d)

Figure 7.4: Scatter plots illustrating the differences and similarities of the interference-aware geocasting algorithm. Results from centre-of-intersection scenarios, all stations transmitting and use the interference-aware geocast algorithm with $X = 10$ or $X = 2$. S is the ratio of the maximum transmission range to required coverage radius. The colour of the points represents the vehicle density configuration the data point is sourced — the units are in vehicles per km per lane.

It is however interesting to note that in Figure 7.4, the data points achieving high PRR are of different colours. In Figure 7.3e (greedy algorithm), only scenarios with vehicle density at 10 vehicles/km/lane could achieve high PRR, whereas Figure 7.4a shows that scenarios with up to 20 vehicles/km/lane (green) is able to achieve high PRR using the interference-aware algorithm (retransmission parameter $X = 10$). Furthermore, reception is improved by suppressing transmission, as demonstrated by Figure 7.4b where X is lowered to 2. Looking at the horizontal positions of the data points of the same colour (same vehicle density), Figure 7.4 indicates that reducing the retransmission parameter (X) reduces the measured load on the channel in most cases.

When the transmission range of each station decreases (Figures 7.4c, 7.4d), necessitating multi-hop forwarding, the benefit of the interference-aware algorithm is not as pronounced. This is because the behaviour of the interference-aware geo-

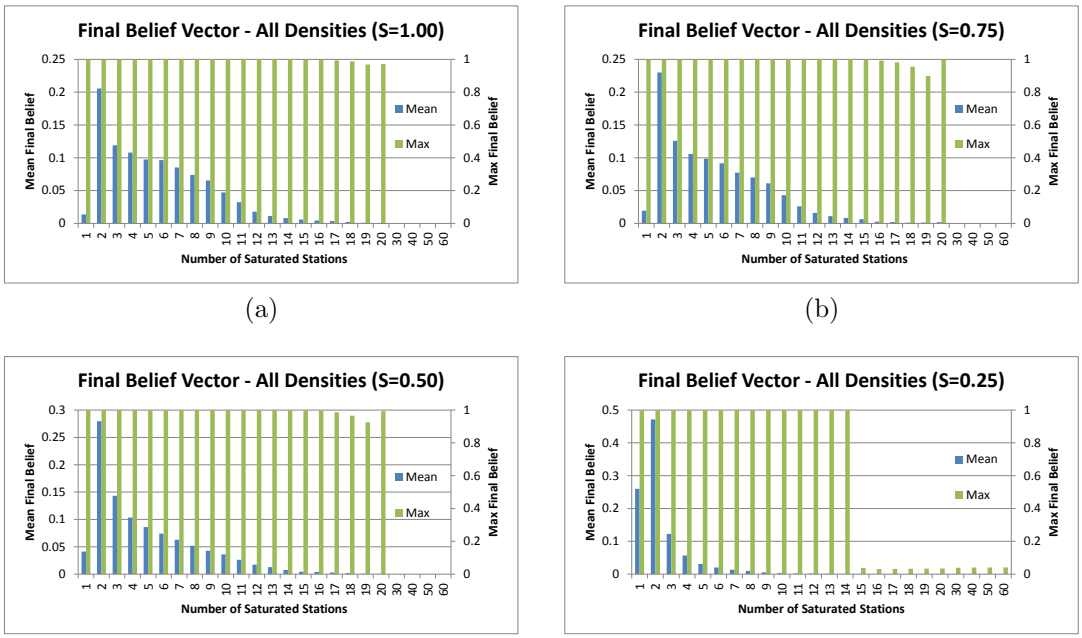


Figure 7.5: The mean and max values of stations' belief vectors. Low max values suggest that perhaps the corresponding bin is unnecessary. Results from centre-of-intersection scenarios, all stations transmitting and use greedy distance-based forwarding algorithm. S is the ratio of the maximum transmission range to required coverage radius. The results show that categories above 20 is rarely used by any estimator, and the mean value for the high contention believes are small.

cast algorithm converges to that of a greedy distance-based algorithm in these scenarios. When this happens, the furthest stations rebroadcast the packet with the parameter X adjusting the distance-based delay, high X shortens the delay. Shorter delays increase contention and can cause a lowering of PRR for dense enough cases, whereas extended delays cause more stations not to rebroadcast, also resulting in a reduction in PRR. Methods of optimising the choice of X for a given vehicle density can improve the performance of the geocast algorithm.

Number of categories for belief vector

Finally, Figure 7.5 shows the mean and the maximum values of stations' final belief vectors. It is observed that at the completion of the simulation running the greedy distance-based forwarding algorithm, bins above 20 are almost never used. The figure also suggests that, at $S = 0.25$, some stations still have a small

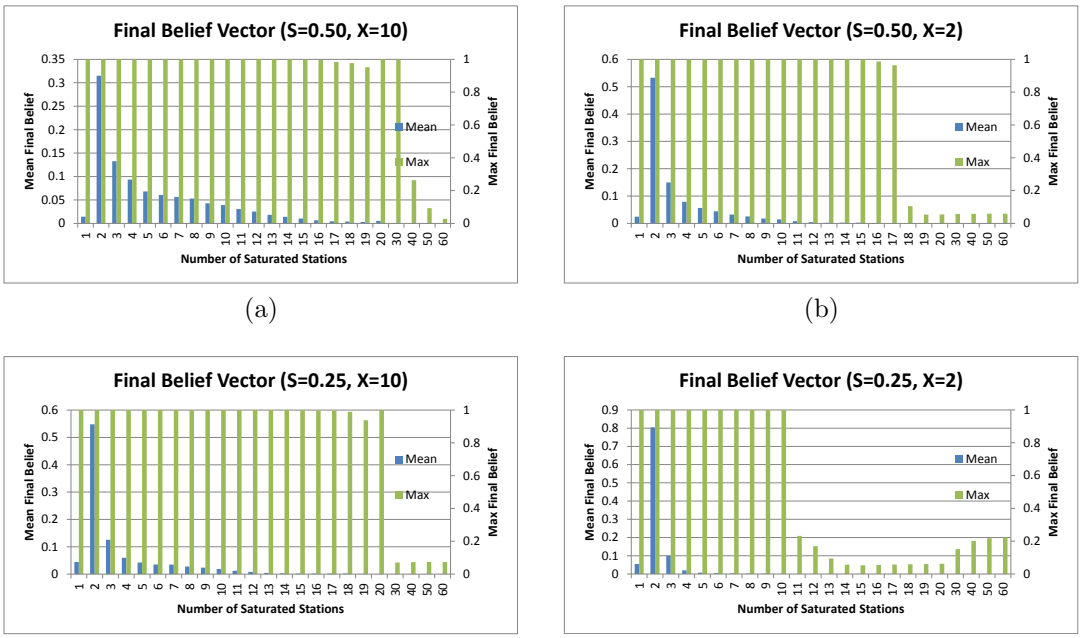


Figure 7.6: The mean and max values of stations' belief vectors. Results from centre-of-intersection scenarios, all stations transmitting and use interference-aware geocast algorithm. S is the ratio of the maximum transmission range to required coverage radius. These results show that while categories above 20 are not used often, a number of stations still retain high estimates. The mean values for the high contention believes are nonetheless quite small.

yet almost-negligible belief in the number of saturated stations being above 15. Stations exhibiting believes in high contention at $S = 0.25$ (which is intuitively a false belief) had only obtained a very small number (< 3) of valid idle period observations (slot count observed is within one contention window), therefore the estimator would not have collected sufficient samples, resulting in their belief vectors having a very wide spread.

Final belief vectors for scenarios running the interference-aware geocast algorithm is shown in Figure 7.6. Figure 7.6a shows that when the channel saturation is high enough, the distribution of belief vector values are similar to those for the greedy distance-based algorithm. An interesting distribution is observed in the not-as-congested scenarios, where a hump can be seen at the tail of the distribution. The actual dataset shows that stations exhibiting similar distribution typically have more than 80% of the inter-frame slot count being higher than one contention window. This suggests that, similar to the observations in the greedy

Algorithm 7.1 Simple algorithm to predict good/bad PRR

```
function ESN_PREDICT(ESN)
    return ESN ≤ ESN_t
end function
```

algorithm, the belief in the high load is likely to be a false high. One possible explanation is that these stations are sparsely connected, leading to bursts of retransmission being observed (high observed contention during bursts, but then discards the longer than CW observations between bursts).

7.2.3 Using ESN to predict/estimate packet reception

This section presents further statistical analysis on the result set, and validates the use of the ESN to predict packet non-reception. Simulation results suggest that the step-like shape of the PRR-vs-ESN relationship (Figures 7.3e and 7.3f) may be useful in predicting whether PRR is acceptable based on the observed ESN. The utility of this behaviour can be assessed by first devising a simple threshold test on the ESN, then collect the basic test statistics of Positive Predictive Value (PPV), Negative Predictive Value (NPV), Sensitivity and Specificity.

For this analysis, the test (Algorithm 7.1) returns positive if a station is predicted to have good PRR (above 80%), and negative otherwise. the test statistics can be interpreted as the following conditional probabilities:

- Positive Predictive Value (PPV) — probability that a station has good PRR if the test returns positive. (*i.e.* trustworthiness of a positive result.)
- Negative Predictive Value (NPV) — probability that a station has poor PRR if the test returns negative. (*i.e.* trustworthiness of a negative result.)
- Sensitivity — probability that a test returns positive given a station has good PRR. (*i.e.* likelihood of correctly identifying a good station)
- Specificity — probability that a test returns negative given a station has poor PRR. (*i.e.* likelihood of correctly identifying a poor station)

Table 7.2: Test statistics of PRR prediction — Greedy distance-based algorithm

S	1.00	0.75	0.50	0.25
PPV	0.16433	0.17887	0.24715	0.17188
NPV	0.99947	0.99374	0.99502	0.97470
Sensitivity	0.98861	0.90372	0.95879	0.94809
Specificity	0.81074	0.78658	0.73832	0.30445
Reward	2.96315	2.86291	2.93929	2.39912

Furthermore, the "optimal" threshold ESN ($ESN_{t_{opt}}$) can be computed by running the test on all data points collected, and executing the following nonlinear program to find the optimum.

$$ESN_{t_{opt}} = \underset{ESN_t}{\arg\max} \quad \sum_S PPV + NPV + Sensitivity + Specificity$$

The nonlinear program is performed using Excel® Solver Add-on running an Evolutionary algorithm on the results obtained from the simulation. The optimal threshold ESN for the greedy distance-based forwarding algorithm is found to be 2. Table 7.2 shows the test statistics using the optimal threshold ESN.

This result suggests that, as predicted, using a threshold ESN value can very accurately determine non-reception of packets (high NPV), but is almost useless in determining whether packets will be received (low PPV). This can be explained by the fact that at high load, the predominant cause of packet loss is packet collision — a high ESN can therefore predict poor PRR. On the other hand, when channel contention is low, the probability of collision due to contention is much reduced, therefore packet loss due to other factors such as shadowing, fast fading and hidden terminals become prominent. ESN does not account for these causes, thus the PPV of this test is low.

The test has high sensitivity because high PRR in the presence of high channel contention is extremely unlikely, making false negatives uncommon. Specificity is also reasonable, but is more likely due to the skewed dataset than the properties of the test — it merely states that in the given dataset, most of the packet loss

Table 7.3: Test statistics of PRR prediction — Interference-aware geocast

	$\%\leq 2$ ESN	$ESN_{t_{opt}}$	Total Reward for ESN_t		
			Optimal	2.0	2.5
Greedy	37.83%	2.000000	11.1645	11.1645	11.1541
$X = 0.5$	98.66%	6.325872	9.22484	8.68678	8.65548
$X = 1.0$	99.36%	5.240879	9.25348	9.20124	9.15830
$X = 2.0$	66.30%	2.039410	10.7448	10.7394	10.7409
$X = 3.0$	56.04%	2.200228	11.1215	11.1159	11.1182
$X = 4.0$	50.44%	2.095224	11.4464	11.4429	11.4383
$X = 5.0$	46.48%	2.000000	11.5421	11.5421	11.5314
$X = 6.0$	44.36%	2.023405	11.6311	11.6370	11.6245
$X = 7.0$	42.45%	2.000000	11.7410	11.7410	11.7306
$X = 8.0$	33.47%	2.276144	12.2306	12.2293	12.2286
$X = 9.0$	39.88%	2.000000	11.8561	11.8561	11.8471
$X = 10.0$	38.80%	2.000000	11.8639	11.8639	11.8536

is caused by channel contention.

Table 7.3 investigates the usefulness of the ESN contention measure for interference-aware geocast. This table details only the total reward value (sum of the four test statistics over the four transmission ranges). One can see that when the retransmission parameter X is high enough, the optimal threshold ESN approaches to the one found for the greedy algorithm. The total reward value is also similar to (and often exceeds) that of the greedy algorithm. Further investigation into the low X cases (which has much lower total reward value) shows that in these cases, while PPV is low and both NPV and sensitivity are high as in the high X cases, there is significant difference in specificity. The specificity is low at low X and high at high X. Together with the results from the low transmission range scenarios in the higher X cases, the highly variable specificity further supports the hypothesis that the observed specificity is simply an artefact of the dataset rather than an intrinsic property of the threshold ESN test.

The implications of these test statistics on the use of ESN for predicting PRR can therefore be summarised in Figure 7.7.

Figure 7.7: Summary of test statistics on PRR prediction using threshold ESN

PPV	**NPV**
Low	0.9–1.0
can't predict success	predicts failure
Sensitivity	**Specificity**
0.9–1.0	Highly variable
finds most successful cases	is load dependent

7.2.4 Channel use effeciency

The analysis also shows that the interference-aware algorithm has a better PPV than the greedy distance-based forwarding algorithm, meaning that the percentage of true positives (stations with ESN ≤ 2 that also has good PRR) is much higher. This can either be due to the interference-aware algorithm generates higher overall load (shifting false negative data points to the right to become true negatives), or that the algorithm is more consistent in improving PRR at lower load (turning false positives into true positives by improving reception).

In order to differentiate between the two possibilities, the PPV from the greedy algorithm can be compared to the interference-aware algorithm with parameter X set to produce the most similar proportion of stations that results in positive tests. Table 7.4 compares the greedy algorithm (37.83% positive tests) to the interference-aware geocast with $X = 10$ (38.80% positive tests), and shows that the interference-aware geocast indeed has a higher PPV than the greedy algorithm even controlling for the proportion of positive tests. This is therefore highly suggestive that the interference-aware geocast is more efficient in generating retransmissions that improves PRR.

Table 7.4: Comparison of PPV values for the greedy distance-based forwarding algorithm and interference-aware geocast, $X = 10$

	Greedy	Interference-aware
$S = 1.00$	0.1643	0.4513
$S = 0.75$	0.1789	0.3990
$S = 0.50$	0.2472	0.3750
$S = 0.25$	0.1719	0.2434

7.2.5 Summary

Through collecting measurements of channel contention using various metrics, it was shown that ESN values produced from idle slot-based contention estimator is able to accurately predict non-reception of packets. In networks with heavy enough contention, the optimal threshold ESN was determined to lie somewhere near 2 ESN, and the reward function decreases slowly between the range $[2, 3)$. Furthermore, the channel estimators were found to need only 20 categories, reducing both the memory and runtime cost of maintaining the channel estimates.

Finally, the interference-aware geocast algorithm shows a higher PPV when using ESN threshold to predict packet reception even after controlling for the proportion of positive tests, implying that the algorithm is more efficient in improving PRR than the greedy algorithm.

7.3 Load-reactive geocast algorithm

Having validated ESN as a useful measure of channel contention that is predictive of packet non-reception, it therefore follows that the ESN can be used to dynamically adjust the retransmission parameter (X) of the interference-aware geocast algorithm. In this section, a simple function that couples the load estimator to the interference-aware geocast algorithm is described, moderating X based on the current channel contention. The performance of this closed-loop system is then evaluated through computer simulations using ns-3.

7.3.1 System overview

This load-reactive retransmission system comprises of two main components. First, the channel sensing and adaptation component that is responsible for observing the radio channel and subsequently adjusting the parameter of the second component; and second the interference-aware geocast component responsible for prioritising stations for rebroadcasting a packet, and is triggered by every cor-

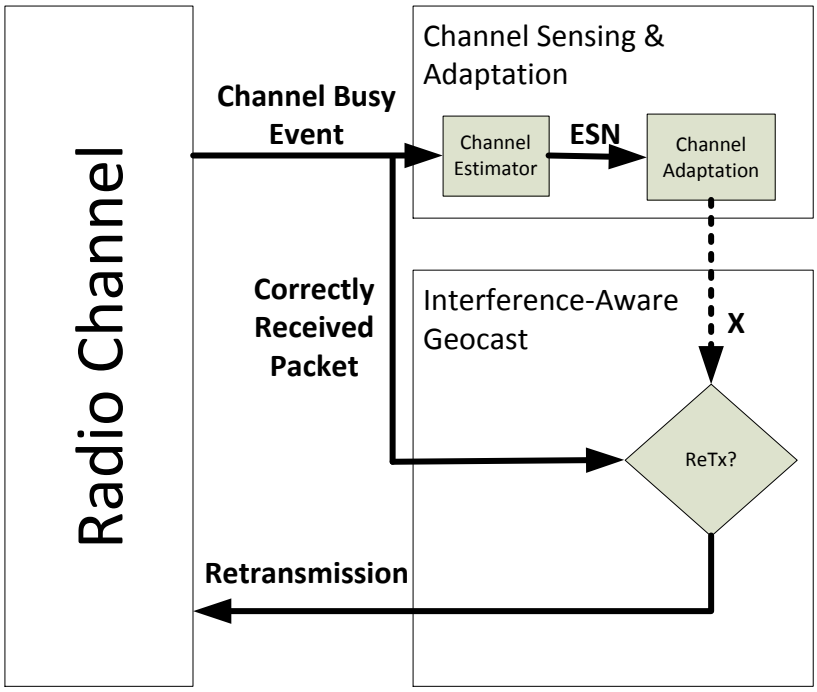


Figure 7.8: The design of the load-reactive system

rectly received packet. The sensing and adaptation component is divided into two parts — a tracking load estimator that is primarily based on the idle slot observation algorithm, and a channel adaptation algorithm that takes the estimated channel condition and compute an appropriate retransmission parameter X for the retransmission algorithm. The channel adaptation algorithm updates the X value every time an inter-frame slot count is available (*i.e.* at the start of every channel busy event), whereas the retransmission algorithm is activated only after a packet has been correctly received before expiry. The retransmission algorithm only inspect the computed X value when it needs to make a retransmission decision. Figure 7.8 depicts the overall architecture of the system.

7.3.2 Tracking channel load

In order to implement this system, the first challenge is to estimate the instantaneous channel contention. The idle slot-based channel estimation technique presented in Chapter 5 uses Bayesian inference, and assumes the channel state does not change by weighing all observations equally. This causes the estimator

to retain infinite memory. When the underlying channel is constantly changing, the output of the estimator will converge to the long-term average and not the instantaneous value. This is undesirable when estimates for the current condition are required. Techniques for enabling tracking include sliding window, Partially Observable Markov Decision Process (POMDP), Hidden Markov Model (HMM) techniques, and attenuating the posterior probabilities before using it as the prior probability of the estimator (which is a special kind of HMM technique).

Sliding window basically involves the estimator keeping track of the N most recent observations, and computing the current condition using only those observations within the sliding window. This requires each station to maintain not only its belief vector, but also a fixed sized buffer, greatly expanding the memory requirement of the simulation. Furthermore, since the operations of the estimator on the belief vector is not reversible (*i.e.* cannot undo the effect of an "expiring" observation), current estimates need to be recomputed from scratch after every observation. This slows the algorithm by a factor proportional to the size of the sliding window.

Partially Observable Markov Decision Processes (POMDP) and Hidden Markov Model (HMM) techniques are alternatives to sliding window, but are also not applicable in general to this problem. HMM is a generalised version of Bayesian inference where stations may move between states (number of estimated saturated stations) with some known probability (represented as a transition matrix). POMDP further extends HMM such that the actions of the station (such as changes in retransmission probability) has known effects on the transmission matrix. In this case, both the transition matrix and the effects of the actions are difficult to ascertain, and are also highly dependent on the number of stations on the network (not modelled in the original derivation). For these reasons, POMDP and HMM techniques in general are not applicable for this purpose.

The last technique is to add an attenuation factor to the previous estimate before using it as the prior probability to the next observation. This limits the

Algorithm 7.2 Algorithm to track network load

```
Let b be the belief vector for ESN.
for all b_i ∈ b do
    b_i ← 1/length(b)
end for

loop
    ô ← observed number of idle slots.
    Define p as the prior probability at the current time step.
    p ← γb + (1−γ)/len(b)                                  ▷ γ ∈ [0, 1]
    denominator ← 0
    for all p_i ∈ p do
        denominator ← denominator + p_i(T_ô|N = i)
    end for
    for all p_i ∈ p do
        b'_i ← p_i(T_ô|N=i) / denominator
    end for
    b ← {b'_i  ∀i}
end loop

ESN_est ← Σ_i i b_i          ▷ Estimated ESN is the weighted sum of belief
```

influence old observations have on the current estimate, allowing the estimator to eventually move to a new value as the underlying condition changes. The choice of the attenuation factor (γ) controls the degree of influence by previous observations — if γ is zero, previous observations has no influence on the current estimate, whereas as γ approaches 1, the effects of the becomes infinite. Technically, this is a special case of HMM methods, where it is assume with probability γ that channel state does not change, and $(1-\gamma)$ that it moves to any state with equal probability. Algorithm 7.2 details the channel estimation algorithm that tracks network load.

By inspecting the simulator traces generated for Section 7.2 it was found that many stations' estimated channel load have converged to within $\pm$1ESN with 95% confidence within 30 observations. For this reason, the attenuation factor (γ) for this experiment was chosen such that the influence of observations from 30 or more steps ago is minimal (attenuated by a factor of at least $\frac{1}{1000}$).

$$
\begin{aligned}
\gamma^{30} &= \frac{1}{1000} \\
\gamma &= 1000^{-\frac{1}{30}} \\
&\sim 0.8
\end{aligned}
\tag{7.1}
$$

7.3.3 Reacting to network saturation

Using the tracking channel estimator, stations can then attempt to influence the interference-aware geocast algorithm, hence channel contention, by adjusting the retransmission parameter X. An increase in X encourages retransmission, hence increases contention; decrease in X suppresses retransmission, hence reduces contention. Section 7.2 shows that there exist a threshold ESN beyond which packet reception is highly unlikely. For this reason, the parameter adjustment algorithm aims to suppress retransmissions when the channel contention is sensed to be above the threshold value (ESN_t).

Furthermore, when X is set to a low value, transmission do not receive the benefits of retransmissions, especially when the network is sparse and needs multi-hop forwarding. The adaptation algorithm therefore not only needs to suppresses retransmissions during heavy channel contention, it also should encourage retransmission when the channel load is sensed to be low.

Finally, when the channel is below saturation, most of the observed interframe idle slot counts are greater than the contention window size. Since the estimator is based on the theoretical model where all stations are saturated and these observations cannot happen, this causes the estimator to produce erroneous results. The algorithm therefore needs to account for this limitation, and encourage retransmissions when the inter-frame idle slot count is too high.

Algorithm 7.3 is designed to achieve the desired load moderation behaviour while accounting for the under-saturation problem by increasing X by a constant when an observation is higher than one contention window and therefore dis-

Algorithm 7.3 Algorithm to react to network load

```
function AdjustX(loadEst, discardedObs, lastX)
    if discardedObs then                          ▷ Too many idle slots
        return lastX + k                          ▷ k is some fixed constant
    end if
    return lastX × ESN_t / loadEst
end function
```

carded. If such parameter increases turn out to be unnecessary, the inflated X would quickly be reduced by the algorithm through its normal operation if the channel load is sensed to be high.

7.3.4 Evaluation

The load-reactive geocast system along with the selected algorithms described in Chapter 4 were evaluated using the network simulator ns-3.9. Parameters used are identical to those in Chapter 4 to ensure consistency, except the more fine-grained set of vehicle densities in Section 7.2 is used.

Section 7.2 shows that the optimal threshold ESN is around 2.0 for busy enough scenarios. For less congested channels, the optimal threshold ESN is higher. The section also shows that, within the range $[2.0, 3.0)$, the reward function is very close to optimal value, but quickly drops when the threshold is outside this range. Therefore in this simulation, the threshold ESN (ESN_t) is set to 2.5, which is within the range $[2.0, 3.0)$ and gives sufficient buffer in case the channel is not very congested. The additive constant (k) is set to 0.75, which appeared to provide a good outcome based on the results from a number of pilot runs using arbitrary values. As will be discussed later, the selection of both ESN_t and k represents a trade off between performance at higher and lower contention situations. Table 7.5 lists the simulation parameters used.

Both the scenarios with only a single tagged transmitter and the scenarios with all stations transmitting are simulated. The single transmitter scenarios are used to determine performance in a known near-optimal (uncongested) situations, whereas the scenarios with all stations regularly transmitting approximate the in-

Table 7.5: Simulation parameters — Load-reactive system performance

Vehicle densities	{10, 15, 20, 25, 30, 40, 50, 60, 75, 90, 100} veh/km/lane
Proportion of HV	10%
Lanes per road	6
Road length	900 m
Intended coverage radius	200 m
Position of source	Centre of intersection 80 m south of intersection Centre of straight road
Antenna Gain (Tx and Rx)	2.512 dB
Rx Threshold	-95 dBm
CS Threshold	-99 dBm
Log-Distant Exponent (γ)	2.0
Log-Distant Ref Loss (at 1 m)	-47.8588 dB (Friis loss, 5.9 GHz)
Nakagami parameter (m)	5.0
Receiver Noise	0 dB
Attenuation across HV	-20 dB
Building Shadow	-30 dB
Max transmission range	{50, 100, 150, 200}m
Transmission rate	6 Mbps
Transmission bandwidth	10 MHz
SIFS	32 μs
Slot Time	13 μs
Packet Rate	10 Hz
Packet lifetime	100 ms
Packet size	54 octets (incl. all headers)
Retransmission parameter X	{0.5, 1, 2, 3, ..., 10}
Number of transmitters	{Single station, All stations}
Additive Constant (K)	0.75
Threshold ESN (ESN_t)	2.5

tended use-case of Cooperative Collision Avoidance systems (CCA). In the CCA use-cases, even though all stations transmit packets, only the packets generated by one tagged station are tracked in the simulation. Background CCA beacons are generated at 10 pkt/s (as proposed for many VANET applications), and in the CCA use-cases, stations are assigned different random start times drawn from a uniform distribution over $[0, 100)$ ms. For each vehicle density setting, 10 situations are tested with results presented being the average of these situations. A sample of 20 tagged packets is taken for each algorithm under test. The performance measure taken is the proportion of packets successfully received before expiry and relevant to the receiving station. Other measures are also recorded but are not relevant to the evaluation. Furthermore, the mean value of X used at each station for determining retransmissions is also logged.

The load-reactive geocast system is compared with the following existing algorithms introduced in Chapter 4:

- No forwarding ("No ReTx")
- Briesemeister *et al.* [84] (greedy furthest successful station — "Furthest")
- Fixed X interference-aware (as presented in Chapter 4), using the best parameter X found. ("Metric")

It should be noted that this load-reactive geocast system would behave significantly differently to the fixed parameter interference-aware geocast algorithm due to the distributed nature of the sensing algorithm. The interference-aware algorithm uses the same parameter X for all stations regardless of the local vehicle density and channel condition, whereas the load-reactive system is expected to cause each station to use a different X. The fixed-parameter interference-aware algorithm can give an indication whether X should increase of decrease as vehicle density changes, but the values of X are not directly comparable.

7.3.5 Results

Reception Performance

Figure 7.9 shows the PRR performance of the load-reactive system (marked as "MetricDyn" in the figure) compared to the other algorithms in the centre-of-intersection scenarios. These scenarios represent the cases where the highest contention occurs closest to the tagged transmitter. The single transmitter scenarios (left hand figures) show that, for busy enough channels, the load-reactive system is able to achieve PRR performance approximately equals that of the best fixed parameter case. (The plots show the PRR for only the best choice of X from the fixed parameter results. The choice of X may be different in each vehicle density setting.) For the not-very-congested channels (10 vehicles/km/lane and where $S = 0.25$), there is an observable difference of approximately 5% between the load-reactive system and the corresponding fixed parameter one, suggesting that the algorithm is not operating as effectively for these low contention scenarios.

The graphs on the right-hand side of Figure 7.9 show the system performance under simulated "real world" load. Here, the difference between the load-reactive system and the others are more pronounced. The results show that the load-reactive system is not aggressive enough in encouraging forwarding at low vehicle densities (density $\leq$ 40 vehicles/km/lane). However, at higher channel contentions, the load-reactive system is sometimes able to outperform the fixed-parameter interference-aware geocast algorithm. When the channel contention becomes very high (*e.g.* $\geq$ 75 vehicles/km/lane at $S = 1.00$), the reactive system once again underperformed. One interpretation of this would be that the load adaptation algorithm is both not aggressive enough at low contention in encouraging retransmission, and does not suppress forwarding enough at very high contention cases. Having said this, the load-reactive system performed at least as well as the no-retransmission case except in one scenario tested (100 vehicles/km/lane). This limitation of the reactive system is also observed for other vehicle layout scenarios (Figure 7.10).

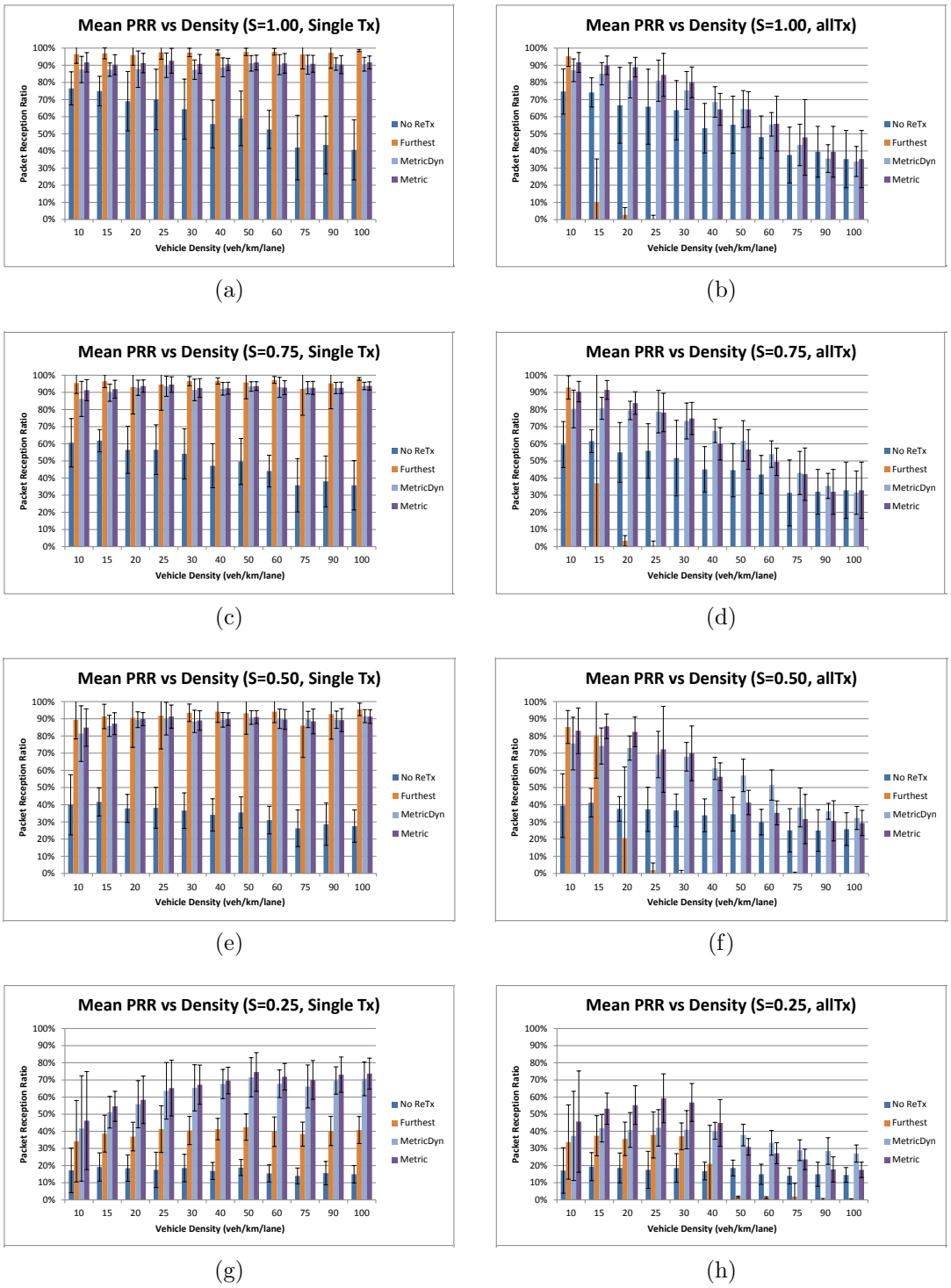


Figure 7.9: Results of centre-of-intersection scenario, error bars represents two standard deviations, $S \in \{1.00, 0.75, 0.50, 0.25\}$ is the ratio of the maximum transmission range to required coverage radius. Parameter X for the metric cases are chosen to give the best PRR. Single transmitter scenarios on the left, all stations transmitting on the right.

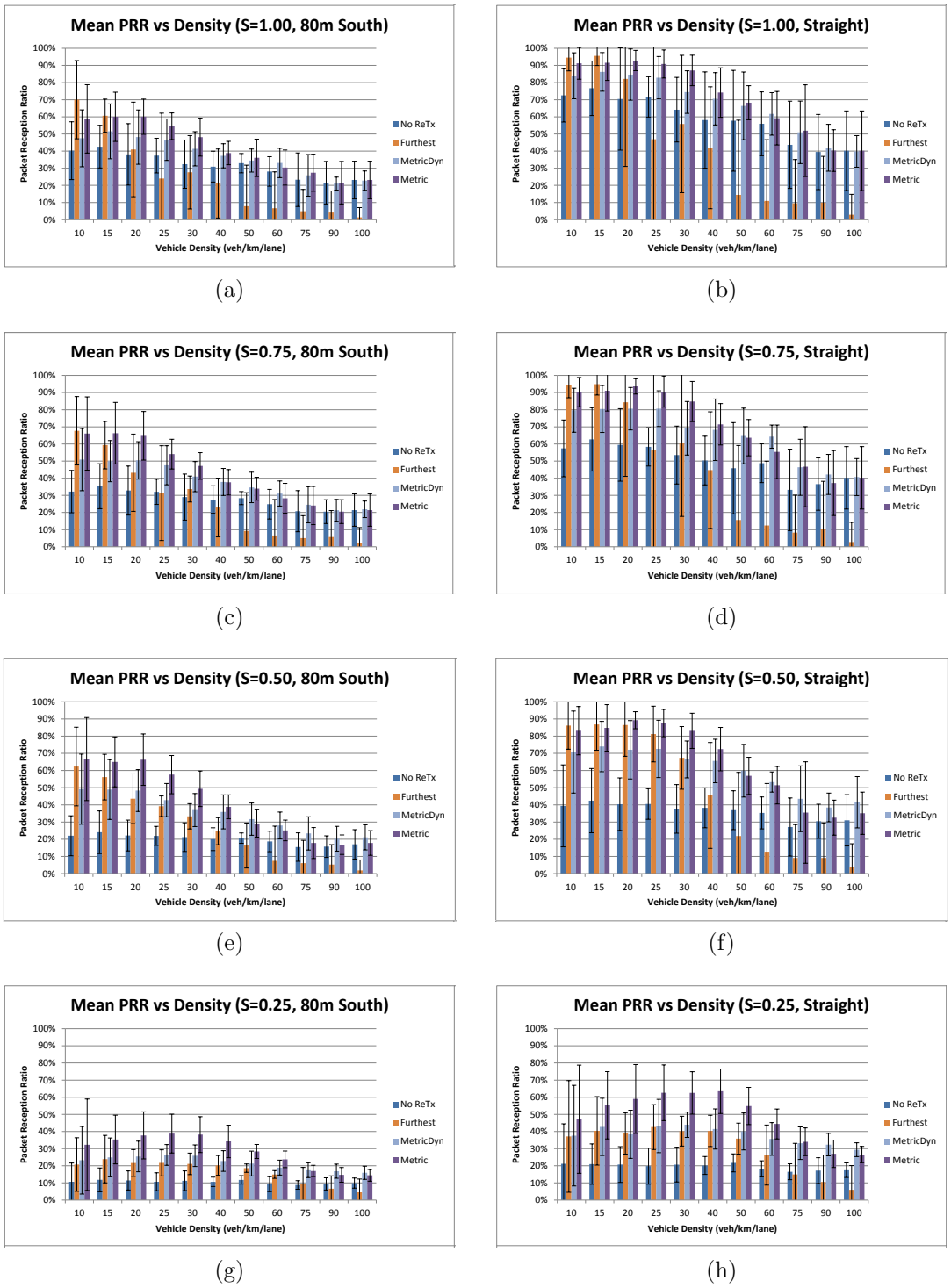


Figure 7.10: Results of the tagged transmitter either 80 m south of intersection scenario or on a straight road, with all stations transmitting, error bars represents two standard deviations, $S \in \{1.00, 0.75, 0.50, 0.25\}$ is the ratio of the maximum transmission range to required coverage radius. Parameter X for the interference-aware algorithm is chosen to give the best PRR for each vehicle density. Intersection scenarios on the left, straight road on the right.

Choice of Retransmission Parameter

In order to investigate how the retransmission parameter is chosen by the load-reactive system, the median X value used is first tabulated, and then compared to the "best" choice of X in the fixed parameter algorithm (Table 7.6). Again, it must be noted that the absolute value of X is not directly comparable between the two algorithms — only the general trend can be compared.

One can see from Table 7.6 that, compared to the algorithm that uses a globally fixed X parameter, the load-reactive system increases X from a rather small value at low vehicle densities before suppressing X again at high vehicle densities. It can be argued that the computed values of X is below the optima at low vehicle densities, while at high vehicle densities, is higher than optimal.

Since X is updated every time the channel ceases to be idle, but the computed value is only used for determining retransmission if the packet was correctly received, the median value of X that is actually used can be different to the computed values. The computed value represents the long-term average of what X should be according to the load adaptation algorithm, while the actual X used would depend on other factors. It is observed that, on average, low vehicle density causes the system to use a value that is more likely to be higher than computed, while high vehicle density triggers the opposite effect.

Figure 7.11 plots the median computed and used X values against the median of measured channel load for each scenario. The plot shows that the majority of the points lie near the threshold ESN, suggesting that the load-reactive system is able to control the load to around EST_t for most cases. The data points that are above ESN_t suggest that, as expected, the system chooses lower values of X as the median load increases in attempt to suppress retransmission for the sensed high load.

Another way to visualise the discrepancy would be to look at the ratio of the actual values of X used for retransmission decision to the computed X. Table 7.7 shows that for low vehicle densities, the actual value of X used tends to be

Table 7.6: Choice of X by reactive and fixed algorithm (Centre-of-intersection)

		PRR		Fixed X	Reactive X (Median)			
	Density	Fixed	Reactive		Computed	Used	Difference	Load
$S = 0.25$	10	45.67%	37.36%	10	4.47	4.77	+0.30	11.54
	15	53.24%	41.80%	10	7.77	7.86	+0.10	7.05
	20	55.28%	40.79%	10	8.88	9.18	+0.30	5.68
	25	59.29%	42.05%	8	9.09	9.00	-0.09	7.45
	30	56.82%	40.94%	5	10.37	10.30	-0.07	3.21
	40	44.89%	40.31%	3	10.66	10.35	-0.31	2.71
	50	30.92%	37.86%	1	7.97	7.50	-0.47	3.14
	60	20.20%	33.30%	1	6.91	6.38	-0.53	2.66
	75	23.59%	28.98%	1	6.53	5.51	-1.02	2.29
	90	17.78%	28.50%	0.5	5.35	4.37	-0.99	2.34
	100	17.53%	27.04%	0.5	4.59	3.58	-1.01	2.39
$S = 0.50$	10	83.00%	75.61%	8	6.18	6.24	+0.06	6.82
	15	85.63%	74.20%	10	9.89	9.52	-0.38	3.10
	20	82.41%	72.94%	4	9.75	9.21	-0.54	3.14
	25	72.26%	69.21%	2	6.12	5.88	-0.24	2.13
	30	69.92%	67.93%	2	6.53	6.33	-0.20	4.63
	40	56.31%	61.17%	1	5.97	5.22	-0.74	5.20
	50	41.31%	57.08%	0.5	3.94	3.20	-0.74	2.54
	60	35.30%	51.46%	0.5	3.51	2.65	-0.86	1.92
	75	31.68%	38.46%	0.5	3.35	2.23	-1.12	1.89
	90	30.63%	36.23%	0.5	2.55	1.68	-0.86	2.20
	100	29.40%	32.31%	0.5	2.29	1.30	-0.98	2.13
$S = 0.75$	10	90.43%	80.33%	10	4.07	4.36	+0.30	11.55
	15	91.42%	80.69%	10	5.88	5.85	-0.03	10.54
	20	83.75%	79.48%	3	6.49	6.51	+0.01	4.62
	25	78.25%	78.79%	2	4.79	4.49	-0.30	5.71
	30	72.77%	73.31%	2	5.04	4.85	-0.19	6.31
	40	59.97%	67.57%	1	5.11	4.20	-0.91	2.61
	50	56.68%	61.69%	0.5	3.05	2.40	-0.65	2.66
	60	49.49%	53.99%	0.5	2.72	1.93	-0.79	2.37
	75	42.24%	42.99%	0.5	3.00	1.74	-1.26	2.02
	90	31.98%	35.43%	0	2.05	1.15	-0.91	1.90
	100	32.91%	31.49%	0	1.86	0.92	-0.94	1.77
$S = 1.00$	10	91.68%	87.07%	10	2.84	3.12	+0.29	13.60
	15	89.99%	85.07%	9	4.46	4.80	+0.34	11.99
	20	88.92%	81.19%	7	5.54	5.63	+0.08	12.59
	25	84.42%	81.00%	3	4.84	4.69	-0.16	10.99
	30	80.03%	75.38%	3	6.36	6.09	-0.26	4.76
	40	64.28%	68.61%	1	6.56	5.56	-1.00	2.58
	50	64.25%	64.45%	1	4.32	3.52	-0.80	2.30
	60	55.92%	55.52%	0.5	4.33	2.88	-1.44	2.39
	75	47.93%	43.52%	0.5	3.96	2.41	-1.55	2.36
	90	39.54%	35.56%	0	2.87	1.61	-1.27	1.83
	100	35.29%	33.86%	0	2.47	1.26	-1.21	1.82

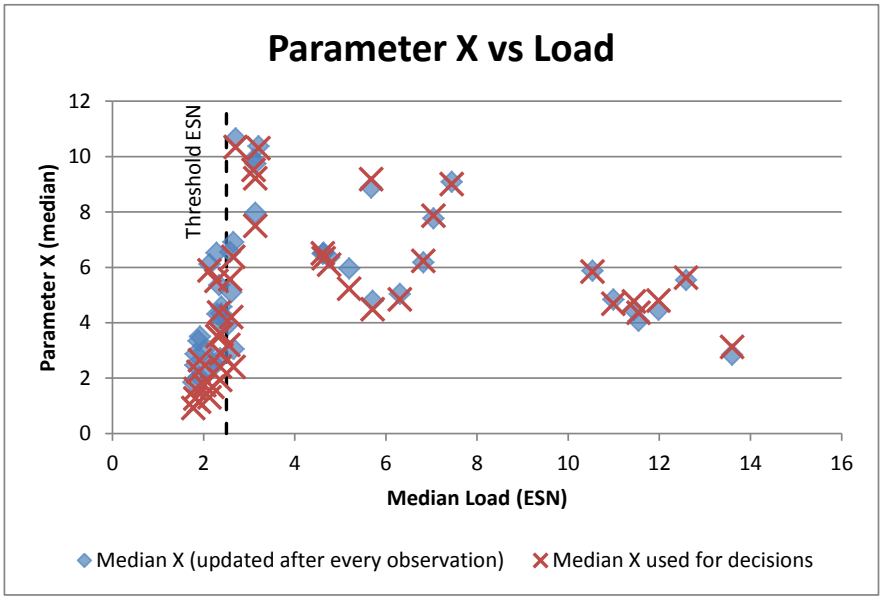


Figure 7.11: Diagram plots the median X value against median load for each scenario of the centre-of-intersection layout. The plot shows that at low load, the actual value of X used for determining forwarding decision is lower than the long-term mean, while at higher measured load, the actual used X value is higher than calculated.

slightly higher than the value computed by the load adaptation algorithm. At high vehicle densities, the values used can be much smaller than those computed by the algorithm. This discrepancy is further discussed in the next section.

Finally, the rightmost column of Table 7.6 contains the median of the final channel load measurement at each station. This value influenced the final choice of X at that station according to the reactive system. Ideally, the load-reactive system should set X such that the final load measure is near 2.5 ESN, the threshold ESN configured. Instead, the final value of load measure is extremely high at low vehicle densities, then almost abruptly drop to a slightly higher-than-desired value before settling at a value around 1.8 to 2.3 ESN.

7.3.6 Discussions and future work

Performance of load adaptation algorithm

Overall, based on the results obtained, one can see that the load-reactive geocast system is able to adapt to the channel load with no manual intervention. However,

Table 7.7: Ratio of actual X used to X computed by the algorithm

	Density	Mean	SD	Min	Q1	Median	Q3	Max
$S = 0.25$	10	1.112	0.255	0.519	1.018	1.104	1.167	4.675
	15	1.064	0.246	0.154	0.950	1.063	1.170	2.455
	20	1.062	0.276	0.063	0.933	1.064	1.177	2.976
	25	1.034	0.300	0.044	0.874	1.026	1.171	3.373
	30	1.053	0.336	0.084	0.873	1.018	1.183	3.176
	40	1.030	0.352	0.052	0.821	1.009	1.197	3.332
	50	0.992	0.344	0.001	0.799	0.974	1.162	3.017
	60	0.977	0.340	0.004	0.791	0.963	1.140	3.546
	75	0.937	0.333	0.003	0.741	0.930	1.109	2.917
	90	0.907	0.357	0.003	0.710	0.899	1.086	4.089
	100	0.873	0.347	0.003	0.675	0.865	1.052	3.760
$S = 0.50$	10	1.059	0.244	0.158	0.947	1.059	1.160	2.926
	15	1.033	0.292	0.197	0.866	1.013	1.161	3.492
	20	1.009	0.354	0.059	0.826	0.995	1.151	4.259
	25	1.013	0.313	0.042	0.844	0.990	1.139	3.541
	30	1.005	0.304	0.070	0.828	0.988	1.145	2.594
	40	0.950	0.317	0.001	0.780	0.944	1.113	2.843
	50	0.895	0.339	0.007	0.699	0.905	1.073	3.076
	60	0.833	0.335	1.3e-4	0.644	0.841	1.030	2.547
	75	0.792	0.362	1.5e-4	0.568	0.801	0.992	3.256
	90	0.754	0.364	1.3e-4	0.520	0.758	0.964	3.806
	100	0.706	0.366	3.3e-4	0.459	0.706	0.924	3.819
$S = 0.75$	10	1.079	0.154	0.452	0.996	1.094	1.168	1.566
	15	1.024	0.228	0.072	0.908	1.032	1.138	1.985
	20	1.031	0.268	0.191	0.897	1.028	1.156	3.209
	25	0.981	0.286	0.051	0.828	0.979	1.114	3.247
	30	0.996	0.285	0.756	0.856	0.991	1.139	3.128
	40	0.900	0.327	0.001	0.730	0.903	1.077	3.255
	50	0.852	0.327	0.003	0.672	0.864	1.028	3.250
	60	0.818	0.337	0.003	0.628	0.830	1.006	3.453
	75	0.728	0.371	1.2e-5	0.478	0.739	0.950	3.200
	90	0.689	0.357	7.4e-6	0.439	0.691	0.910	3.319
	100	0.630	0.361	4.5e-6	0.364	0.622	0.854	2.871
$S = 1.00$	10	1.105	0.133	0.449	1.051	1.113	1.168	1.959
	15	1.057	0.200	0.197	0.977	1.069	1.144	2.851
	20	1.028	0.225	0.268	0.900	1.028	1.143	2.490
	25	1.022	0.231	0.249	0.896	1.017	1.134	2.498
	30	0.988	0.265	0.058	0.846	0.994	1.123	2.533
	40	0.940	0.308	0.017	0.784	0.942	1.100	3.943
	50	0.873	0.318	0.002	0.713	0.892	1.045	4.377
	60	0.815	0.338	0.001	0.628	0.842	1.006	2.286
	75	0.745	0.364	2.5e-4	0.497	0.757	0.973	2.899
	90	0.686	0.366	1.7e-4	0.419	0.685	0.923	2.728
	100	0.654	0.358	1.8e-4	0.393	0.655	0.879	3.172

the load adaptation algorithm is not aggressive enough in encouraging rebroadcast in low vehicle density situations while also not aggressive enough in suppressing retransmissions at high vehicle densities.

It is noted that the parameters of the load adaptation algorithm had not been completely optimised. The load adaptation algorithm uses three parameters — an additive constant (k) that encourages retransmission at low contention, a target threshold ESN (ESN_t), and the function has implicitly an exponent of 1.

The additive constant (k) is required in order to overcome the limitation of the theoretical model the algorithm is based on. The underlying theoretical model for estimating channel contention assumes all stations are saturated (always have something to send). This assumption allows solutions to the theoretical model to be computed in reasonable time — relaxing that assumption requires an prohibitively long computation time, with high inaccuracies caused by floating point precision (see Chapter 6). Unfortunately, this assumption implied that all inter-frame idle periods would be at most one contention window long, which is not the case in this simulation.

The additive constant (k) applies if an observed inter-frame period is over one contention window. In this situation, one would assume that the channel is below saturation, which means retransmissions can be promoted (*i.e.* increase X). In order to increase the aggressiveness in promoting retransmission at low saturations, multiplicative factors don't react fast enough, hence a constant k is used.

The other branch of the algorithm aims to suppress retransmissions at high load. Here, a factor inversely proportional to the sensed load ($ESN_t/load$) is multiplied to the last parameter value to try and push the sensed load to a predetermined threshold (ESN_t) with ESN_t having already been optimised in the first section of this Chapter. The distributed nature of this algorithm means that this measure only controls the current station. By assuming homogeneity, one reasons that other stations in the proximity also would have a similar estimate of

channel load, and thus act similarly, resulting in a reduction of channel contention as a whole. This factor also encourages retransmission at low load, however, an additive constant is more effective because it can increase X quicker when X is very low. To improve the responsiveness of this load-suppression aspect of the algorithm, one could raise the multiplicative factor to a higher exponent. However, increasing the exponent would decrease system stability by causing potentially very large variation in the X value at each step.

Currently, the channel estimation algorithm discards observations that are over one contention window long. This artificially inflates the estimated channel load especially at low contention scenarios. The behaviour can be observed in Table 7.6 where the low vehicle density scenarios returns a very high estimated load. This plays havoc on the multiplicative factor, causing the algorithm to retard retransmission instead of promoting them (evident from the high-load tail of Figure 7.11 — the depressed X does not appear to actually reduce the final sensed load). This effect is partially overcome by the additive constant, but it appears that the constant by itself is not currently enough.

The additive constant is a factor that would always limit the performance of the algorithm. Even though a high k would make the algorithm more responsive at low load, it will cause the algorithm to not retard retransmissions as aggressively at high load. This is because even in very high load, unsaturated stations means that there is a non-zero probability of an inter-frame idle period being longer than one contention window. Hidden terminals further compound the problem.

Optimising the additive constant and the factor exponent would simply be trading off between the high-load and the low-load performance as well as system stability. Higher constant improve low-load performance but reduces high-load performance.

Discrepancies between computed and used X

Simulation results also shows a discrepancy between the value of X used for making retransmission decisions and the mean value of X computed by the algorithm. In order to understand this behaviour, it is important to remember the load-reactive geocast system is actually two separate components coupled together.

The load sensing and adaptation component takes channel measurements as an input, and calculates a control variable (X) that is most appropriate at a specific point in time. The time-average mean computed X value therefore represents the value of X the algorithm had determined to be most appropriate for that station. Looking at this component statistically and assuming steady state, one can view this component as a random process that generates an output distribution with a certain mean.

The second component of this system is the geocasting algorithm. In this component, valid and unexpired packets that had been received correctly are prioritised for retransmission. When such packets arrive, this algorithm inspects the computed X value at that point in time, and then makes a retransmission decision. One can view this as a process that samples the load sensing and adaptation component. In the steady state, both the computed and used X would be optimal, thus would not alter the channel condition. In the simulation, retransmissions are not always triggered — station distances, hidden terminals, packet expiry and other random factors can all cause packets to not be received correctly or in time. The mean value of the "used" X is therefore a sample mean of the load sensing and adaptation output distribution.

The difference between the sample mean and the distribution mean is called sampling error, and is typically caused by two main factors — sample size and sample bias. When sampling a probability distribution, it is likely that the sample mean will be different to the distribution mean. As the sample size increases, the expected sample mean should approach the distribution mean. The sampling size

problem refers to the situation where insufficient number of samples was taken, increasing the likelihood of having large discrepancies between the two means. Sampling bias refers to the dependence between how samples are obtained and the actual value being measured. In this context, the sample size problem implies that the computed X has significant outliers that was by chance not used for retransmission decision making; the bias problem implies that certain computed X causes the retransmission algorithm not to be triggered.

First, the possible sample size problem is considered, and it is found that the discrepancy cannot be fully explained by random sampling error. The number of retransmission a station makes must be the lower bound of the number of times that the station has looked up the computed X value. This is because for every retransmission that stations made, the retransmission algorithm must have looked up the computed X value in order to make the decision. Using that as a lower bound, the maximum standard error of the sample mean for any given confidence interval (*e.g.* 99% CI) can be determined.

The test uses the claim that the discrepancy can be explained by random sampling error (the sample mean is within the standard error of the true distribution mean) as the null hypothesis (H_0). This null hypothesis is rejected if the normalised difference between the true distribution mean (which is known) is greater than the calculated standard error.

Computing the standard error for every station within the target area of each scenario simulated results in 156,110 out of the 236,804 stations (66%) rejecting this null hypothesis. It is therefore reasonable to conclude that sampling error alone is insufficient to explain the discrepancy between the mean computed X value and the mean used X value.

Next, the potential dependence between the triggering of the retransmission algorithm and the computed X is explored, and thereby exploring the insight this discrepancy brings. As previously discussed, the retransmission may not always be triggered after every non-idle period. Non-idle states include when a station

is transmitting, receiving, as well as has sensed a carrier on the channel but the signal's preamble was not detected. Of these states, retransmission decision can only be triggered by the receiving state, and only if the packet was correctly decoded. A sampling bias here means that the system preferentially sample at certain X, which occurs if the correct reception of a single packet (original or rebroadcasted, and from any source) as a ratio of non-busy states is influenced by the retransmission parameter X.

Suppose the mean computed X is higher than the optimal value, *i.e.* the computed value will cause too much contention. Since the instantaneous X values is dependent on the last observed inter-frame idle period (which is random), the individual computed X values will be distributed around the mean. Consider the case where the X value used is even higher than the mean — the channel contention will be worse, resulting in an increased probability of packet collision (*i.e.* packet non-reception but carrier present). Retransmission decision cannot be triggered by a packet not being correctly received, therefore the retransmission algorithm is less likely to sample X when X is too high. On the other hand, if the instantaneous X value is lower than the mean (*i.e.* closer to the optimal), collision is less likely, thus the retransmission algorithm is more likely to be triggered. This creates a bias towards the X values that are closer to the true optima when the mean computed X is higher than optimal.

On the other hand, the reason for the system self-correcting at low contention (mean computed X is lower than optimal) needs to be better understood. The results seem to indicate that the system indeed does use an X value that is higher than the average computed value. However, similar arguments to the higher than optimal case cannot be constructed. Increasing X towards the optimal value increases redundancies, which increases the probability that a station will eventually receive the message. The act of retransmission increases both the number of correctly received packet and total number of non-busy events for stations that is close enough to the retransmitter (thus increases the proportion

of samples), but for those further away such that they can only sense the carrier but is unable to decode, it increases only the total number of non-busy events (*i.e.* reduces the proportion). This is insufficient to conclusively argue for sampling bias that favours the more optimal X without also making assumption on the ratio of stations that can receive the packet to the those that senses carrier only.

Therefore, due to the interaction with the environment, the retransmission algorithm can self-correct suboptimal retransmission parameter from the load sensing and adaptation component if the computed parameter is too high. Furthermore, the system is also observed to self-correct when the computed parameter too low. This is evident from Table 7.7 where the mean and the median value of X used tends to be in the direction of the optimum value (at low contention, the ratio is greater 1, whereas at high contention, the value tends to be much smaller than 1). The cause of the preference when the computed X is lower than optimal still needs to be better understood.

Future work

As stated, the dependence between the retransmission parameter and packet error rate, especially at the low contention levels, needs to be better understood. A better understanding of this behaviour will allow the load adaptation component to better decide the retransmission parameter, utilising not just the instantaneous channel load estimates, but also some measure of the optimality of the last chosen X value.

7.4 Conclusion

In this chapter, one design of load-reactive geocast system that combined the load estimation and retransmission techniques presented in previous chapters is investigated.

The existing metric, channel busy time, is found to be incapable of predicting whether a packet can or cannot be received, whereas the ESN metric is able to

predict with high accuracy packet non-reception. The test to predict packet non-reception is a simple comparison to a threshold value, and despite its simplicity, is shown to be highly accurate in predicting non-reception as well as being able to identify most of the cases that has high reception ratio.

Compared to the greedy distance-based forwarding algorithm, the interference-aware geocast algorithm has more efficient channel use. This inference is based on the better positive predictive value of this test on the interference-aware geocast algorithm.

A design of a load-reactive geocast system that coupled the output of the load sensing algorithm to the geocast algorithm is then presented, allowing the geocast algorithm to adapt to the channel load in order to optimise packet reception. Through computer simulation, the geocast algorithm is shown to be effective in moderating channel load. It was identified that a trade-off needs to be made in relation to system performance under high or low channel load. An emergent behaviour was observed where the two-part system, to a certain extent, self-corrects sub-optimal output from the load adaptation component.

In the next chapter, the issue of simulator accuracy in wireless network research is raised. First, it is shown that simulation outcomes differ greatly amongst the various well known and commonly used simulator packages, none of which conforms to theoretical predictions. Looking in depth into one specific simulator, the nature of the misbehaviour is identified, leading to a set of workarounds that allows the outcomes from that simulator to be interpreted and used correctly.

Overview

This chapter investigates the validity of using existing computer simulation packages for broadcast-mode communications.

Contributions

- **I have identified high discrepancies between outputs of different commonly-used network simulator packages.** The discrepancies amongst the simulators are likely to be caused by errors in the implementation of the IEEE 802.11 MAC-layer broadcast behaviour.

- **I have evaluated the impact of ns-3 broadcast-mode misbehaviour.** By comparing the simulation results to theoretical predictions, I have shown that the misbehaviour observed from ns-3 simulations of broadcast mode IEEE 802.11 transmissions has a small impact in terms of application-layer performance, but has major effects on algorithms that rely on MAC-layer observations such as collision probabilities and idle slot counts. The observations and analysis is applicable to all versions of ns-3 at least from ns-3.4 to ns-3.15. (It is most probable that the workarounds are also applicable to releases after ns-3.0.4 when the YANS Wifi model [1] is first introduced.)

Chapter 8

Variability between Network Simulators

8.1 Introduction

This chapter investigates the accuracy and validity of network simulation packages for broadcast mode IEEE 802.11 communications. Compared to other methods of validating and evaluating network algorithms and theories, such as real-world experiments and experimental testbeds, computer simulations are inexpensive and very flexible. This results in the extensive use of computer simulations in networking research, hence their accuracy is critical.

There are many different network simulators available, ranging from complex general purpose simulators such as OMNet++, ns-2 and ns-3, to more specific simulation engines including JiST/SWAN, to highly specific simulations that are typically developed specifically for a single experiment by the experimenter such as the simple DCF simulation in Chapters 5 and 6. During the course of the work in this thesis, many simulators were tried, and it was found that the results from these simulators differ greatly at times.

First, the MAC-layer outcomes from a range of commonly used network simulation packages are compared to theoretical predictions, showing huge variances

exist amongst simulators, and each showing different non standard compliant behaviours. Next, the discrepancies observed in ns-3.9 are investigated, focusing on how the discrepancies affect simulation outcomes. The analysis observed in ns-3.9 should be applicable to all versions of ns-3 from ns-3.4 (and is most likely to be also applicable to version from ns-3.0.4, *i.e.* when the IEEE 802.11 model was first introduced into the simulator) and up to ns-3.15. A set of "workarounds" that allows users of ns-3 to correctly interpret and use the results from the simulator notwithstanding its non-standard behaviour is also discussed.

8.2 Comparing network simulators

To compare the behaviours of the different simulators in the broadcast context, the same network scenario is simulated using each simulator separately. For a range of saturated station counts, the distribution of idle slots observed on the channel is compared to the theoretical predictions. Where possible, confounding physical layer "enhancements" implemented by the simulators (*e.g.* packet capture effect, shadowing and fading, hidden terminals) are turned off in order to investigate only the MAC layer implementation.

Networks containing some specified number of saturated stations is simulated. Each station on the network is kept saturated either by a loop-back that causes a frame to be added to the transmit queue when a frame is sent by the station, or by higher layer queuing more packets for transmission than the capacity of the channel allows. The frames transmitted are all broadcast mode frames. The following subsections describe the implementations and configurations of each simulator used.

8.2.1 Simple DCF model

The simple DCF model used in Chapter 5 is used here as the baseline to compare the various models with. The model implements the relevant backoff mechanism

for broadcast model exactly as specified in IEEE 802.11 standard, abstracting out all timing and physical layer effects. The model simulates a simplified DCF backoff counter that merely decrements at each time slot and reset when reaching zero. This model simulates the following:

- Fixed size contention window (CW) for each station.
- backoff counter reinitialise to a uniformly distributed value within the CW after transmission by the station. This models the DCF broadcast behaviour (*i.e.* no ACKs) and assumes all stations are saturated.
- Global (shared) timeline in "slots". Data transmission, IFS, *etc.* occur between slots and the actual wall time for the action is ignored.
- Transmission is lost if and only if there is a collision (two or more stations scheduled to transmit in the same slot)
- Model assumes all stations are synchronised (propagation and processing times are zero and no hidden stations). Without assuming synchronisation, the time between slots cannot be ignored as stations that are not synchronised will see different slot boundaries.

When this model is executed, each station simulated is assigned a random backoff counter value uniformly distributed over the contention window. At each time step, all backoff counters are decremented by one if the counter value is greater than zero. If the counter is zero, it is assumed that the station will initiate a transmission, and the counter is reset to a backoff counter uniformly distributed over the contention window. The transmission is assumed to be successful if only one station initiated a transmission, and assumed to have failed due to collision if more than one station transmitted. If no station initiated a transmission at that timeslot, then the channel is considered idle at that time, otherwise, the channel is considered busy. In this simulation, statistics on idle periods, probability of channel being busy and packet success ratio are collected. Further information on this model is located in Chapter 5.

8.2.2 OMNet++ INET model

OMNet++ offers two main general-purpose wireless communication modules: INET and MiXiM. Only the INET module was investigated in this study. In this simulation, the standard `Ieee80211Mac` model in the INET module is extended to provide extra feedback without changing its original behaviour. Specifically, extra feedback is added to the MAC model to notify the application layer when transmissions are completed via a side channel and not as a MAC-layer control signal. Extra logging facility is added to the MAC model to count the frequency of the various observed interframe idle slot counts.

In addition to the extension to the MAC-layer model, a new MAC-layer packet generator is implemented. This packet generator simulates a saturated station by subscribing to the transmission completion event (generated by the extended MAC layer) and pushing a new MAC SDU onto the MAC layer as soon as the previous frame had completed transmission. This packet generator removes the need to use a constant-bit-rate packet source that generates more packets than the channel can support (thus overflowing the transmit queue). This greatly reduced memory use and the size of the log file as the packets are no longer being dropped at the MAC transmit queue.

Each scenario is run 10 times using different seeds for the pseudo-random number generator. A 50 m $\times$ 50 m area with the specified number of non-moving stations is simulated. All stations simulated are within reception range of each other (no hidden stations) and uses the packet generator to generate broadcast MAC frames (destination address "`ff:ff:ff:ff:ff:ff`"). The simulation is allowed to warm up for one simulated second before data is collected for 60 simulated seconds. Table 8.1 summarises the parameters used.

Two variants of station layouts are tested due to complications resulting from these non-realistic scenarios. First variant — all stations are uniformly distributed in the field. In these tests, the distances between stations introduced propagation delays, affecting the accuracy of idle slot observations. Second variant — all

Table 8.1: Simulation parameters — OMNet++

Parameter	Value
repeat	10
sim-time-limit	60 s
warmup-period	1 s
*.playgroundSizeX	50
*.playgroundSizeY	50
**.radio.bitrate	2 Mbps
**.radio.transmitterPower	2.0 mW
**.radio.thermalNoise	-110 dBm
**.radio.sensitivity	-85 mW
**.radio.pathLossAlpha	2
**.radio.snirThreshold	4 dB
*.channelcontrol.carrierFrequency	2.4 GHz
*.channelcontrol.sat	-110 dBm
*.channelcontrol.alpha	2
**.mac.maxQueueSize	14
**.mac.rtsThresholdBytes	3000 B
**.mac.bitrate	2 Mbps
**.mac.retryLimit	7
**.host[*].mobilityType	NullMobility
**.trafficGen.destAddress	ff:ff:ff:ff:ff:ff
**.trafficGen.startTime	0 s
**.trafficGen.waitTime	0.0005 s
**.cwMinBroadcast	[3, 7, 15, 31, 63, 127, 255]
Lan80211.numHosts	up to 450

stations are placed in exactly the same position. In these tests, the idle slot observations observed falls directly on slot boundaries (as expected), and thus gives an accurate measurement of idle slot distribution. However, in the second set of tests, the INET model is unable to calculate the receive power correctly (division by zero error), causing all packet reception measurements to be registered as zero. For data analysis, the idle slot observation results from the second set of station layouts, and packet reception results from the first set of layouts are used.

In order to determine the interframe idle slots, the time between a station switching from any busy state (`CCA_BUSY`, `TX`, `RX`) to `IDLE`, and the same station switching from `IDLE` to any busy state are recorded. This time is then converted into the number of backoff slots by first subtracting DIFS and then dividing the difference by the slot time.

Only the trace from the first station in each experiment is used when aggregating the results. This is because all other traces are identical as all stations are within range of each other and stochastic propagation loss (*e.g.* fast fading) is not used.

It is important to note that the packet capture effect cannot be turned off. Therefore, the packet reception ratio observed may be higher than those observed from the simple DCF model even though the idle slot counts is expected to mirror the outcomes from the equivalently configured simple DCF model.

8.2.3 Ns-2.34 CMU model

The second simulator investigated is the ns-2 simulator. Version 2.34 of ns-2 and above provides a number of different wireless simulation models, of which two relates to ad-hoc simulations: "`Mac/80211`" (from Carnegie Mellon University, referred to as the "CMU model" in this chapter) and "`Mac/80211Ext`" (the newer model from Mercedes-Benz Research and Development North America and Karlsruhe University). The official documentation from ns-2 currently recommends that `Mac/80211Ext` be used in place of the old CMU model, but an abnormality observed when using the new model makes the use of the CMU model necessary.

Initially, a wireless simulation model using `Mac/80211Ext` was constructed. However, when trying to determine the various constants to use to calculate idle slot count, it was observed that some values collected from the simulation does not make sense. The set of times observed between transmissions when there is only 1 active transmitter is completely different to those when there are multiple transmitters, with the time difference not being the sum of any integer multiple of PIFS and slot times. This suggests that the MAC layer calculation of the transmission time for the transmitting station is longer than the PHY layer's calculation. Assuming that PHY layer does not report the completion of broadcast transmission to the MAC layer, this discrepancy could have caused the MAC layer of the transmitting station to wait for longer than the actual transmission time

before starting the DCF process. This error would not affect non-transmitting stations as they rely on the PHY layer to report the finishing of the packet reception. This potential bug practically reduces the total number of saturated stations in the network as the stations that were previously transmitting may have its backoff counter desynchronised from the rest of the network. Therefore the old CMU model is used to construct the simulation instead. As an aside, this bug also means that the outcomes from the `Mac/80211Ext` simulation cannot be compared to the others as the number of idle slots cannot be determined.

The simulation constructed using the CMU model involved creating simple extensions in TCL script. A simple constant-bit-rate traffic generator is implemented on top of an UDP/IP stack, with the generator sending broadcast datagrams. The generator produces packets faster than the station could transmit, creating a transmit queue that is always full (hence a saturated station). In addition, a flooding message passing agent is also attached to each station to ensure saturation. Each scenario is run 10 times using different seeds for the pseudo-random number generator. An area of 10 m $\times$ 10 m is simulated, with all stations positioned on the same spot (5 m, 5 m, 0 m), ensuring that all stations are within range of each other and the propagation time is zero. One extra observing (non-transmitting) station is placed to collect channel statistics. Each simulation is run for 60 simulated seconds.

Data from these ns-2 simulations are harvested from the log files generated by the simulator. Only the entries from the observing station corresponding to packet reception and collision are processed. A collision event generates one entry in the log for each packet involved — one entry at the conclusion of the first packet involved in the collision, and one entry at the beginning of other transmissions in the collision. Since there are no hidden terminals in this simulation, collision occurs only if stations select the same backoff slot for transmission. Because all transmissions are of the same duration, the packet collision entries in the log mark the beginning and the end of a transmission. The packet reception entry

Table 8.2: Simulation parameters — Ns-2

Parameter	Value
Channel type	`Channel/WirelessChannel`
Physical Layer	`Phy/WirelessPhy`
Propagation model	`Propagation/TwoRayGround`
Antenna type	`Antenna/OmniAntenna`
MAC type	`Mac/802_11`
Interface Queue type	`Queue/DropTail/PriQueue`
Interface Queue length	50
Maximum contention window	[3, 7, 15, 31, 63, 127, 255]
`Mac/802_11 basicRate_`	3 Mb
`Mac/802_11 dataRate_`	3 Mb
Packet size	200
Traffic generator period	0.0005 s
`Lan80211.numHosts`	up to 450

is recorded when the packet had been completely received (end of transmission). Table 8.2 details the simulation parameters.

Using the log files, statistics on the time difference between the end of a transmission and the beginning of another (either explicitly indicated by a collision entry or implicitly by subtracting the frame duration from the packet reception entry) are collected. The number of interframe idle slots is calculated by first subtracting DIFS from the times recorded and then dividing the difference by slot time. If the remaining time after subtracting DIFS is not an integer multiple of slot time, the EIFS (instead of DIFS) is subtracted from the time recorded before the division. The validity of subtracting EIFS is confirmed by the fact that the interframe period that is not an integer multiple of slot time happens only after a collision entry where the IEEE 802.11 standards specify EIFS to be used.

8.2.4 Ns-3.9 WiFi model

The third simulator investigated is the ns-3 simulator. Ns-3 is a newer version of the network simulator, designed to replace ns-2. It comes with one standard implementation of the IEEE 802.11 communication stack. The IEEE 802.11 implementation in ns-3 is redesigned from ns-2 with the internal mechanics simplified. The sim-

Table 8.3: Simulation Configurations — Ns-3.9

Parameter	Value
WiFi PHY model	`ns3::YansWifiPhy`
Mobility model	`ns3::ConstantPositionMobilityModel`
stations' positions	(0, 0, 0)
Channel model	`ns3::YansWifiChannel`
Propagation loss model	`ns3::RangePropagationLossModel` (unit disc)
Unit disc radius	9×10^9
Propagation delay model	`ns3::ConstantSpeedPropagationDelayModel`
WiFi MAC model	`ns3::NqosWifiMac`
WiFi protocol	`WIFI_PHY_STANDARD_80211a`
EIFS - DIFS	0 s
AIFSN	2
Maximum contention window	[3, 7, 15, 31, 63, 127, 255]
WiFi rate adaptation	`ns3::ConstantRateWifiManager`
Network socket model	`ns3::PacketSocket`
Application model	`ns3::OnOffApplication`
Application On/Off Time	1 s/0 s (*i.e.* Always on)
Application data rate	60 Mbps
Packet size	800 bytes
Number of IFS observed	50,000

ulation is known to produce valid results when simulating unicast packets [177], but there is a known (and long-standing) bug [186] that affects broadcast mode transmissions. This investigation looked at ns-3.9's behaviour, including the effects of the bug. Since the internal mechanics of the IEEE 802.11 implementation had not change since at least ns-3.4 (and possible since ns-3.0.4) until the bug was partially fixed in ns-3.16, the observations here is applicable to all releases between ns-3.4 and ns-3.15.

In the ns-3 simulations, the required number of saturated stations are created as specified. An extra observer station is added to the simulation in order to determine packet reception ratio. In these simulations, all stations are positioned in the same location $(0, 0, 0)$ using `ListPositionAllocator`. All stations are set to be static (non-moving). The saturated stations are installed with `PacketSockets`, with an `OnOffApplication` attached. The application is configured to be always on, generating constant bit rate traffic at 60 MBps, divided into 800-byte packets. This configuration is well above the maximum throughput supported by the

underlying layers, thereby creating a saturated station. Furthermore, all stations are configured such that the timing for EIFS is the same as DIFS. The simulated channel is a 20 MHz-wide channel at 5.9 GHz, with unit disc propagation loss (`RangePropagationLossModel`) and a maximum range of 9×10^9 metres. This is sufficient to ensure all stations are within range of each other. Data rate adaptation on the stations are switched off by setting the data rate management function to `ConstantRateWifiManager`.

Table 8.3 shows the configurations for these simulations. Parameters not listed uses their default values.

8.3 Results

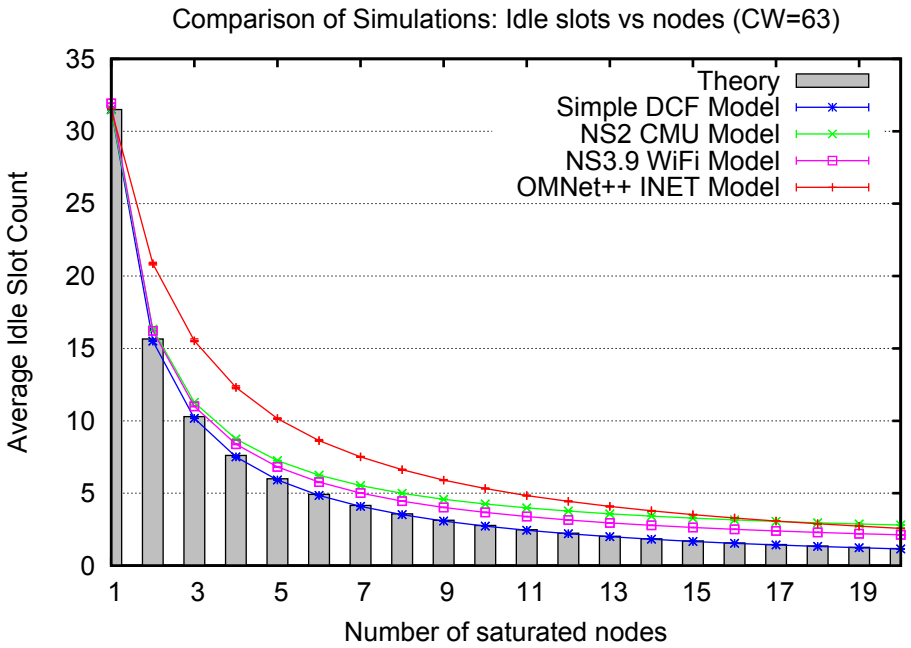
Similar to Chapter 5, statistics collected from these simulators are the overall interframe idle slots counts, packet error ratio, and the distribution of observed idle slot counts. Figure 8.1 plots the statistics observed from the simulations (columns represents the predicted values from theoretical model).

Comparing the lines in Figure 8.1 visually, one can see that all simulators show similar trends as the number of saturated stations increases. Of the four simulators tested, the simple DCF model shows the closest match with the theory. This is due to the fact that the assumptions used in the model matched the assumptions used in the theoretical model closely. Once "real world" effects are added into the simulation (as in the other simulators), the results start to deviate.

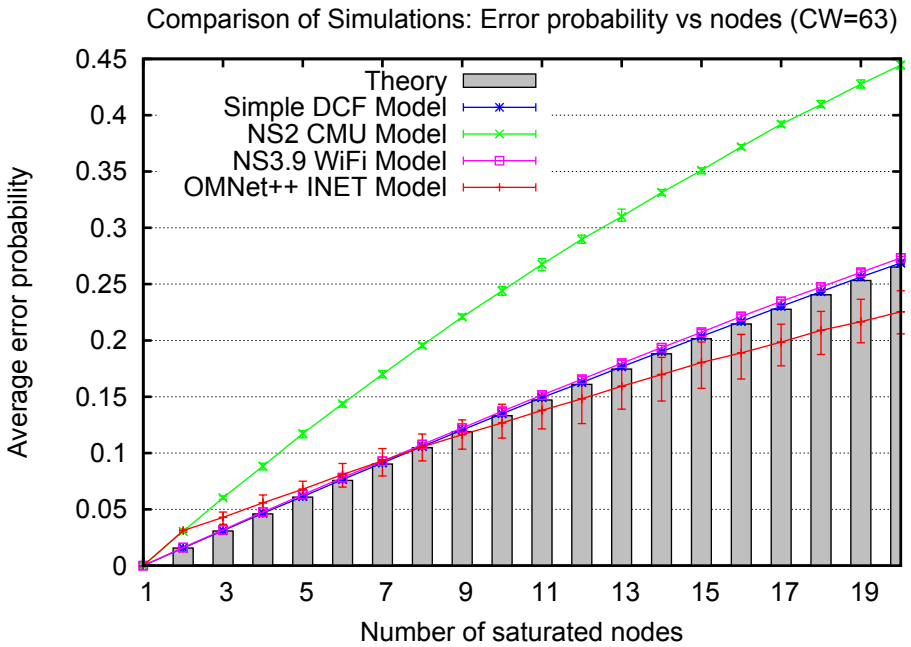
OMNet++ simulations produce result curves that do not decay as fast as the model predicted, even when all stations are positioned in the same location. This higher spread in idle slots contributed to the lower packet error rate than predicted.

The result curves from ns-2 simulations match the idle slot predictions closer than the OMNet++ simulations. However, the ns-2 CMU model produces much higher packet error probability than predicted.

Of the three complex network simulators tested, ns-3 produces results that



(a) Mean idle slots across different simulators



(b) Mean packet collisions across different simulators

Figure 8.1: Overall network statistics for a contention window size of 64 as a function of the total number of concurrent saturated stations, in the absence of hidden terminals as observed from simulations of OMNet++ INET model, ns-2 CMU model, ns-3 WiFi Model, and the simple DCF model, compared to theoretical predictions.

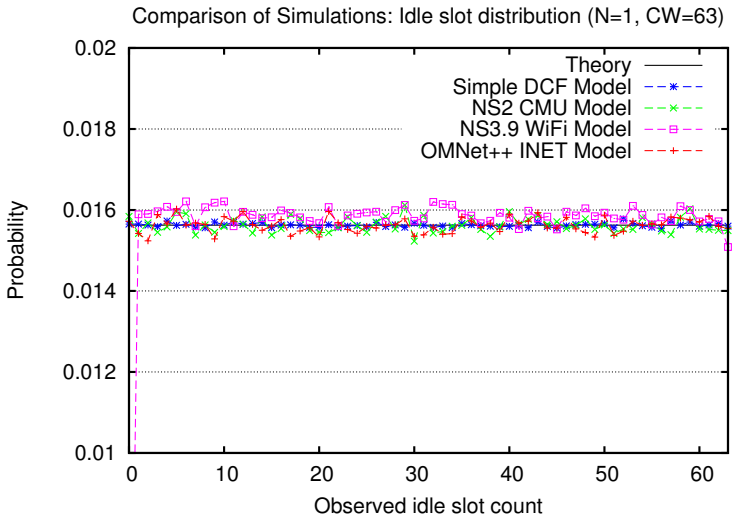
Table 8.4: Goodness-of-fit for overall statistics across different simulators, CW=63, comparing to theoretical predictions

		DCF	ns-2	ns-3	OMNet++
Idle Slots	R^2	0.99993	0.93623	0.98290	0.86798
	RMSE	0.00310	1.48866	0.05886	2.48288
Packet Error	R^2	0.99986	0.80895	0.99934	0.65578
	RMSE	0.00020	0.13124	0.00036	0.11247

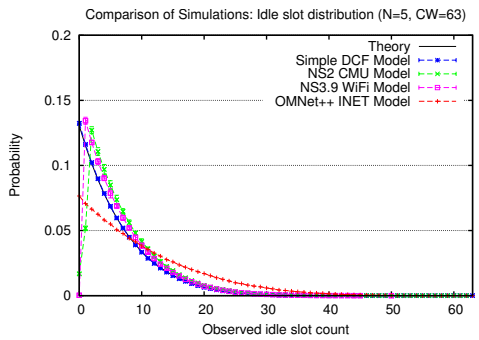
match the predicted values most closely, notwithstanding the known bug in the simulator. The simulator produces idle slot observations that are fairly consistently one higher than expected, while the packet error rate is almost as expected.

Table 8.4 calculates goodness-of-fit values for these simulators, quantifying how well the theoretical model fits the output from these simulators. As expected, the theoretical model shows high correlation with the simple DCF model output in terms of both mean idle slots and error probabilities. The model also fits ns-3 output fairly well, registering both a high R^2 value and low RMS error. Both OMNet++ and ns-2 output deviates from the predicted values, and do not resemble each other, demonstrating the high variability amongst the different network simulators. Figure 8.2 further investigates the reason for the discrepancy, plotting the observed distribution of idle slots (dashed lines) and compares them to theoretical predictions (solid line).

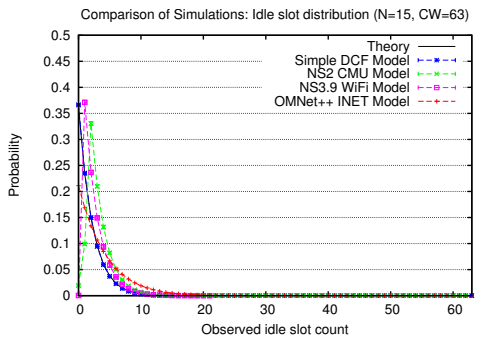
Looking at the distribution of observed idle slots, the high variability amongst network simulators are obvious. Notably, both ns-2 and ns-3 generate a much lower than expected probability for immediate transmission (*i.e.* no backoff after DIFS is rare), whereas OMNet++ produces a distribution that has a much higher spread. For both ns-2 and ns-3, other than their lower than expected probability for immediate transmission, their plots match the predictions fairly closely. Figures 8.2d and 8.2e suggest that ns-3 shifts the plot to the right by one slot. This behaviour can be explained by the known bug [186]. The causes of ns-2 and OMNet++'s discrepancies are unknown. Table 8.5 shows the goodness-of-fit values for these plots. It should however be noted that both the R^2 values and RMS errors for ns-2 and ns-3 had been heavily skewed by the anomaly at low slot



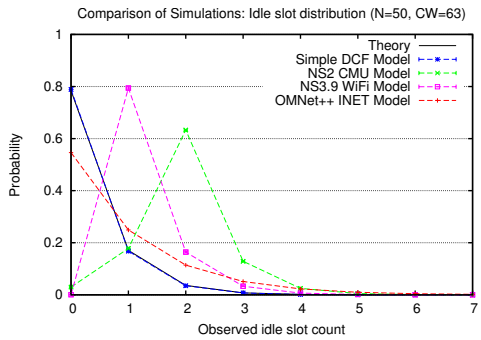
(a) 1 station (y-axis range from 0.01 to 0.02 only). ($\frac{1}{64} = 0.015625$)



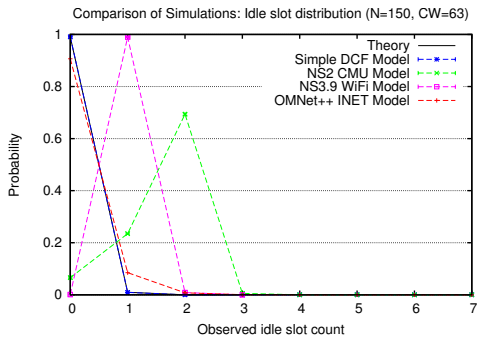
(b) 5 stations



(c) 15 stations



(d) 50 stations



(e) 150 stations

Figure 8.2: Distribution of idle slots between transmissions for a contention window size of 64, as observed from simulations of OMNet++ INET model, ns-2 CMU model, ns-3 WiFi Model and the simple DCF model, compared to theoretical values from Chapter 5. All simulation results are aggregated over 10 executions of the simulation using different random seeds. Error bars denote two standard deviations from the mean of the values observed, and most are too small to be visible. Error bars omitted for (a) for sake of clarity — values observed ranges between 0.014 and 0.018 for both ns-3.9 and ns-2 models, and between 0.015 and 0.016 for the simple DCF model.

Table 8.5: Goodness-of-fit for idle slot distribution across different simulators, CW=63, comparing to theoretical predictions

		DCF	ns-2	ns-3	OMNet++
1	R^2	1.4×10^{-11}	0	0	0
	RMSE	6.7×10^{-6}	0.00081	7.9×10^{-5}	0.00018
5	R^2	0.99984	0.62176	0.65292	0.64537
	RMSE	1.5×10^{-5}	0.01867	0.00097	0.01273
15	R^2	0.99989	< 0	0.10529	0.63557
	RMSE	2.4×10^{-5}	0.07956	0.00627	0.03263
50	R^2	0.99994	< 0	< 0	0.75065
	RMSE	3.0×10^{-5}	0.29223	0.03943	0.08048
150	R^2	1.00000	< 0	< 0	0.97805
	RMSE	4.7×10^{-6}	0.58857	0.11009	0.05638

counts.

It is also observed that the OMNet++ INET module does not trigger EIFS backoff after a packet collision (making the implementation standards non-compliant, however this observation is irrelevant to this study).

8.4 Discussion

Based on the results obtained, it is obvious that the simulators' broadcast behaviours all deviate from the specifications in the standards. Even though it is acknowledged that the additional "real world effect" such as fading, packet capture, *etc.* can explain some of the deviations observed, the procedures adopted in Section 8.3 should have minimised these effects.

In terms of the results from OMNet++ simulation, the plots obtained form a reasonable curve that is never-the-less different to the expected behaviour. The extra spread in the idle slot distribution (lower maximum probabilities and longer tail), explains the higher than expected mean idle slot count and lower error probability.

On the other hand, the ns-2 mean idle slot count curve matches the expected values closer than OMNet++ at low load, but subsequently deviates from the expected values further than OMNet++. This behaviour can be explained by the

idle slots distribution plots observed — ns-2 results, for some unknown reasons, shifted the peak from 0 idle slots (which is expected), to 2 slots, with the remainder of the distribution quickly fall back towards the expected values. This heavy concentration of distribution at 2 slots also partially explains the poor reception ratio — higher contention at slot 2. However, the observed error probability is still too high for this to be the sole cause.

The third simulator, ns-3 produces curves that are most consistent with the expected behaviour. The anomaly observed in the output is consistent with the long-standing bug [186]. This bug has the effect of rescheduling almost all broadcast packets that are scheduled to be transmitted at the slot after DIFS (*i.e.* those frames in the transmit queue with a backoff counter of 0). Effectively, this bug manifests in the idle slot behaviour such that it is rare for packets to be transmitted with no backoff, and the backoff counter is mostly uniformly distributed between $[1, aCW_{min}]$. Other than this rescheduling (which happens without delay) ns-3 exhibits the correct behaviour. This off-by-one behaviour is observable in Figure 8.2, and explains the slightly higher than normal packet error ratio (slightly higher channel contention due to the unavailability of slot 0), and the slightly higher mean idle slot count (basically off by one).

Besides the behaviour of the simulators, the ease of customising and discovering the simulation parameters is also important. The ease of discovering and changing these parameters is vital to these discussions as the discrepancies may have been caused by misconfiguration. Simulators whose parameters are difficult to find and/or change would lead to easier misconfiguration. In the process of conducting this study, ns-2 appeared to be the most difficult to configure, with some settings needing to be set in multiple parameters. However settings are not explicitly stated in available easily accessible documentation, therefore it is easy to change one setting but not changing another. Definitive listing of parameters and associated documentation for the ns-2 CMU model is not easy to find, even though there are many tutorials on the internet.

On the other hand, both OMNet++ and ns-3 are fairly easy to discover parameters and are relatively easy to configure. OMNet++ provides a graphical (tabular) listing of all relevant parameters, and are easy to modify. However, values of some of the parameters are not straightforward to retreieve during the simulation. The newest simulator, ns-3, does not provide a graphical editor for configurations. However, a comprehensive listing of the parameters is readily available from the documentations (assuming the model authors follows programming guidelines). Setting and retrieving model parameters programmatically in runtime as well as in the form of configuration files are straightforward once the verbose syntax to do so is understood. (The API documentation provides a listing of all possible variations of the syntax.) On the flip side, ns-3 misconfiguration are often not easy to discover during runtime and debugging is often difficult as there are many ways to modify the parameters, with each parameter also have many aliases.

8.5 Further characterisation of ns-3 behaviour

In order to gain better understand of the behaviour of the ns-3 WiFi model, to assess the impact of the immediate access bug [186] and to ascertain the validity of broadcast mode WiFi simulations in ns-3, further measurements are made using ns-3. A more in-depth understanding of this bug allows a set of workarounds to be derived, which not only allows retrospective analysis of previous results, but also to guide future use of this simulator. In short, the ns-3 bug bogusly generates a "collision event" during broadcast mode operations (where no end-of-transmission event is implemented) when a packet is queued for transmission with no contention backoff while another concurrent broadcast mode transmission has just completed (in simulation-time, but the end-of-transmission status is not yet been processed). The end result in a single saturated station scenario is that the idle slots count observed is almost uniformly distributed over the range $[1, aCW_{min}]$ instead of $[0, aCW_{min}]$ as specified in the standards. Bug investiga-

tion results reported in the bug tracker [186] suggest that the packets bogusly collided are not dropped, but are resubmitted at the head of the transmit queue, resulting in slightly lower throughput due to the extra backoff, but with no extra packet loss.

Even though the ns-3 community had investigated the bug, the main emphasis of their investigation was to reduce the number of bogusly reported "virtual collisions" on individual stations only. The effects of this bug on how stations interact was not investigated. Conceptually, the bug could manifest in one of the following ways — (i) observations follows the distribution predicted if all stations individually chooses a backoff counter in the range $[1, aCW_{min}]$ (*i.e.* bug affects stations individually and does not affect how stations interacts); or (ii) observations follows the distribution predicted if all stations chooses a counter value in the range $[0, aCW_{min} - 1]$, then right-shift by one (*i.e.* bug affects station interactions globally, and the individual station behaviour observed is a consequence of the global interaction misbehaviour).

For this investigation, the simple DCF model (and a modified version) is used as a proxy for the "correct behaviour" to be compare to the ns-3 simulation. The simple DCF code is used in preference to the theoretical model because the theoretical model is computationally much more expensive than the simple DCF model, and the previous section had shown that the model is very accurate. The simple DCF models (and modifications) tested are — (i) modified DCF model with same size contention window such that stations never choose slot 0 (corresponds to behaviour (i)), (ii) normal DCF model with contention window size one smaller than simulation parameters then manually right-shift the result (corresponds to behaviour (ii)), and (iii) normal DCF model with no adjustments (the standard behaviour).

8.5.1 Results

The idle slot count and packet error probabilities are collected from the ns-3 simulations using a range of contention window sizes, and then compared to the expected results based on the candidate models.

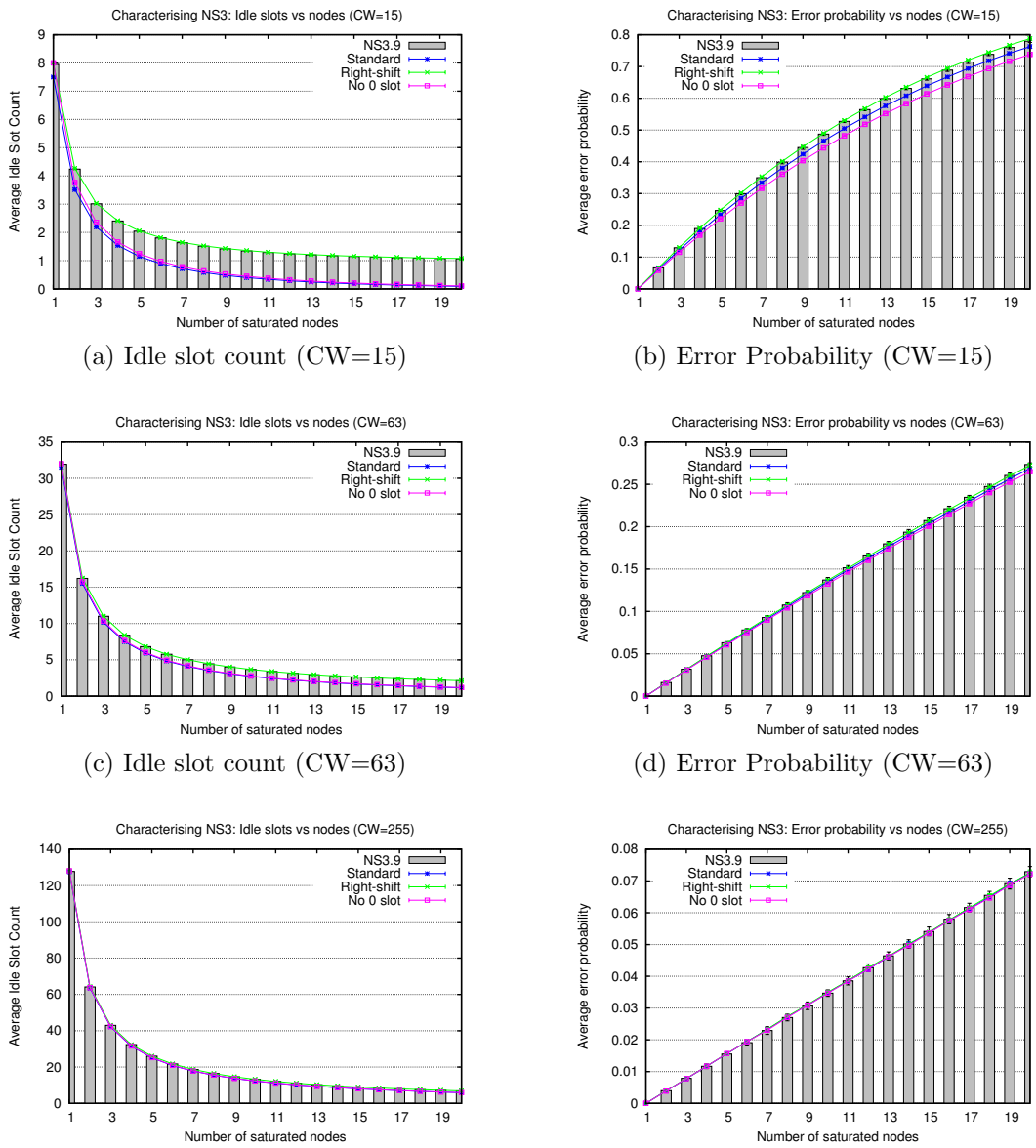
Figure 8.3 shows that the model where all stations are affected globally (all stations adopts a contention window of one smaller than specified, and right-shifts the chosen slot by one) most closely matches the observations. This model accurately predicts the expected average idle slot counts, and only slightly over-estimates the error probability at higher channel load. Figure 8.4 further investigates the expected behaviour based on this model in terms of the distribution of observed idle slots.

The observed idle slot distributions suggest that the right-shift model can accurately predict the observed distribution, even though it does underestimate slightly the probability of immediate DCF transmissions. As the channel load increases, the probability of observing immediate DCF transmission in the faulty implementation decreases (*i.e.* follows the right-shift model more closely), improving the accuracy of the adjusted model.

8.5.2 Discussion and potential workaround

These results show that even though a known bug exists in ns-3's implementation of the IEEE 802.11 broadcast behaviour, the simulation model is still useful for broadcast communication research like VANETs. When the contention window used is large enough, the effect of the bug in terms of the overall statistics diminishes as expected. The off-by-one error when the contention window is large enough would only increase the overall network contention slightly. Since the error rate experienced and the extra delay of one slot is almost negligible, the overall throughput predicted by the model should be accurate enough for most research.

In terms of the algorithms that rely on MAC-layer statistics such as the ob-



(a) Idle slot count (CW=15)

(b) Error Probability (CW=15)

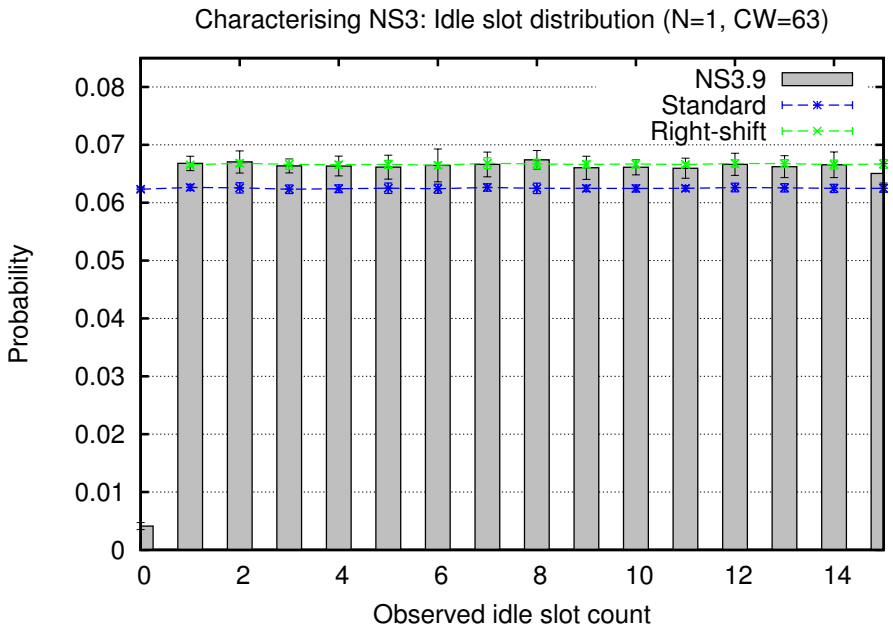
(c) Idle slot count (CW=63)

(d) Error Probability (CW=63)

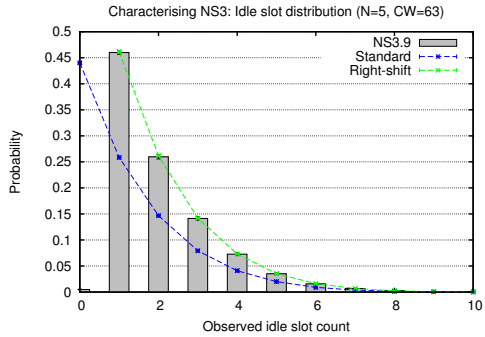
(e) Idle slot count (CW=255)

(f) Error Probability (CW=255)

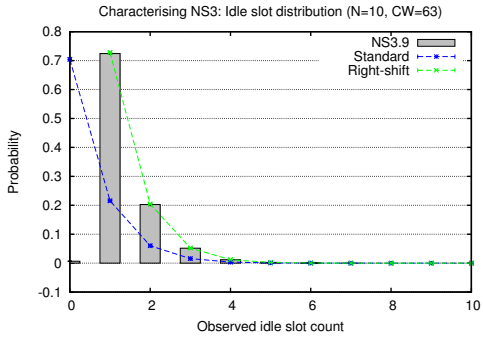
Figure 8.3: Overall statistics from ns-3 simulations. Comparing various theoretical models (estimated from equivalent simple DCF simulations) to ns-3 observations. All simulation results are aggregated over 10 executions of the simulation using different random seeds. Error bars represent two standard deviations from the mean values.



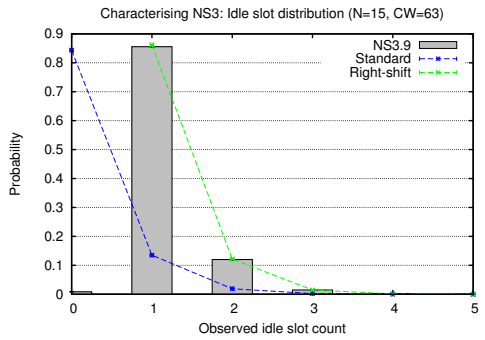
(a) 1 station



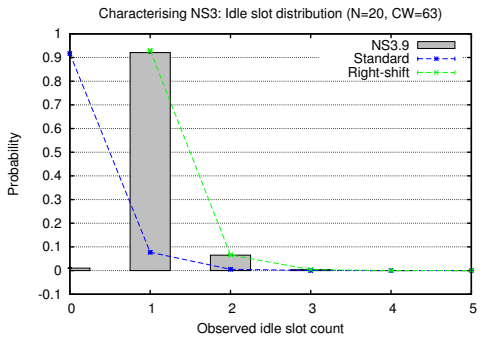
(b) 5 stations



(c) 10 stations



(d) 15 stations



(e) 20 stations

Figure 8.4: Observed and predicted distribution of idle slots between transmissions for a contention window size of 15. Theoretical predictions computed using the simulation of the simplified DCF model of contention window size 15, with results right-shifted by one. All simulation results are aggregated over 10 executions of the simulation using different random seeds. Error bars represent two standard deviations from the mean.

servation of idle slots (*e.g.* Idle Sense [141], both Bianchi and Tinnirello's [136] and the idle slot based channel contention sensing methods), implementations of these algorithms need to be slightly modified to account for the bug. Alterations to the existing methods would be very minor because the bug only manifests as a right-shifted idle slot distribution of a slightly smaller contention window size. As a workaround for these algorithms relying on MAC-layer statistics, the algorithms and/or implementations should be changed to:

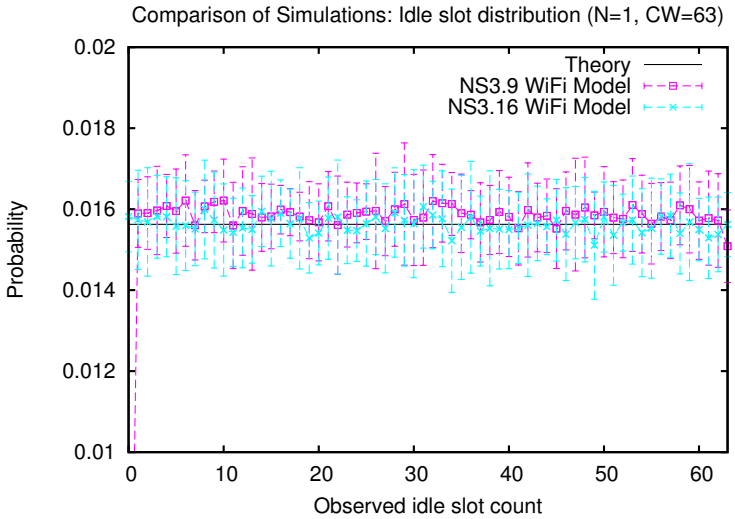
1. Reduce the contention window size by 1 in the theoretical analysis (or increase the simulated contention window size by 1).

2. Increase DIFS by 1 slot time in the analysis (or subtract 1 slot time from the interframe idle period before interpreting the results).

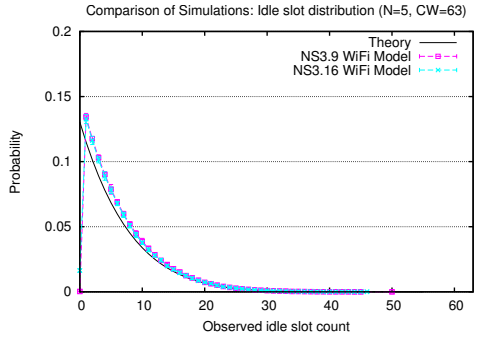
For example, to simulate the operation of Idle Sense for $aCW_{min} = 15$, one would increase the "optimum idle slot count" by one in the algorithm (workaround 1), and run set aCW_{min} to 16 in the simulation (workaround 2). Similarly, for Bianchi and Tinnirello's sensing algorithm to work almost as expected, one would sense the channel busy status at slot 1 (*i.e.* skip one slot time after DIFS before continuing with the algorithm), which can be achieved by the sensing algorithm using an AIFSN of 3 (instead of 2, which is DIFS). Finally, the contention detection work in Chapter 5 would also operate correctly (and thus use ns-3 to test the algorithm) if the scheme is configured for a contention window size of one less than specified for the simulation, and uses an AIFSN of 3 instead of 2.

8.6 A note on ns-3.16

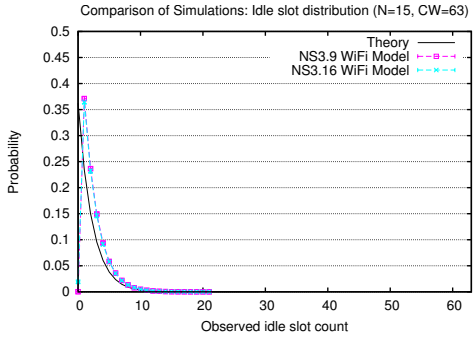
The patch that supposedly fixed the "Immediate DCF Access" issue had been incorporated into the ns-3.16 release. The patch implemented explicit end-of-transmission event for broadcast transmissions, which should resolve the problem regarding the bogus "virtual collisions". To investigate the correctness of the



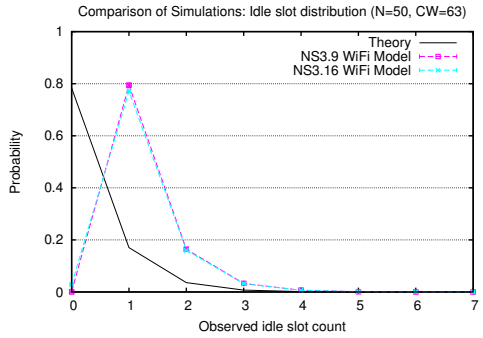
(a) 1 station (y-axis range from 0.01 to 0.02 only). ($\frac{1}{64} = 0.015625$)



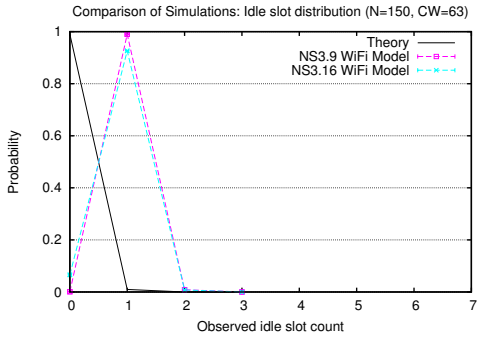
(b) 5 stations



(c) 15 stations



(d) 50 stations



(e) 150 stations

Figure 8.5: Distribution of interframe idle slots for a contention window size of 64, as observed from simulations of ns-3.16 and ns-3.9 WiFi Models, compared to theoretical values from Chapter 5. All simulation results are aggregated over 10 executions of the simulation using different random seeds. Error bars denote two standard deviations from the mean.

new simulation model and the applicability of the workarounds, test cases used in ns-3.9 are ported into ns-3.16 and then run.

Results from the tests conducted show that ns-3.16 still suffers from problems with its DCF behaviour, and the new behaviour is different in nature to the one in ns-3.9. Figure 8.5a shows that when only a single saturated station is on the network, the station behaves as expected, with interframe idle slot counts uniformly distributed across the contention window. However, when the number of saturated stations is increased, the behaviour differs from the theoretical expectations. Figure 8.5 plots the distribution of idle slot counts for varying numbers of saturated stations. It shows that the observed distribution sits somewhere between those observed from ns-3.9 and their respective theoretical predictions. The probability that slot 0 is used is non-zero in ns-3.16, increases to a maximum at slot 1 with the maximum value between the theoretical expectation and the observed value in ns-3.9, and then subsequently decays as expected. Quick inspection of these results suggests that ns-3.16 may have corrected the individual station behaviour, but the interaction between stations remains incorrect.

8.7 Conclusion

In this chapter, the IEEE 802.11 broadcast behaviour is investigated in three commonly used network simulator packages (ns-2, ns-3 and OMNet++). All three simulator packages exhibit behaviours that deviate from the IEEE 802.11 specifications, and each deviates in a different way. OMNet++ INET model produces a wider spread of observed idle slots than expected from theory, and subsequently results in lower packet error probability than expected for a given channel load. Ns-2 CMU model right-shifts idle slot distribution by two slots, but no corresponding bug report is found at the time of writing. Ns-2 also generates a higher-than-expected packet error rate, even if the unusual idle slot distribution had been taken into account. Ns-3 contains a known bug that shifts the idle slot

distribution to the right by one slot, but otherwise produces a result that closely matches the theoretical expectations.

The abnormal behaviour of ns-3 was characterised further, and one can see that the results from ns-3 broadcast mode simulations are plausible notwithstanding the known bug in the system. For simulations that investigate only the large-scale effects, the simulation results are fairly accurate if a large enough contention window (or low enough channel contention) is used. For simulations requiring MAC-layer statistics, a two-part workaround that allows ns-3 results be correctly interpreted was presented.

Based on the simulation outcomes obtained for ns-3.16, it is advisable that, until the causes and consequences of the new misbehaviour are understood, this version (and ns-3.17, which contains no changes to the IEEE 802.11 implementation) not be used for simulations that rely heavily on MAC layer statistics. For simulations where MAC layer statistics is not critical, any version could be used as long as the contention window is set large enough.

Chapter 9

Conclusion and Future Work

This thesis presented works aimed to improve packet reception for broadcast messages in vehicular ad hoc networks. It first considered the inherent channel congestion problem in VANET multi-hop geocasts by presenting an efficient cooperative retransmission algorithm. The algorithm takes into account the amount of redundant or irrelevant packets received by vehicles both within and outside the intended target area, in addition to the amount of additional coverage the relay offers. Using computer simulations that also take into account attenuation caused by obstructions along the line-of-sight between vehicles, it is shown that interference-aware algorithm is effective in improving reception of multi-hop geocasts. Results show that the interference-aware geocast algorithm can outperform other geocast algorithms while exposes a parameter can be used by another algorithm to further control the amount of packet redundancy the algorithm generates. The algorithm's efficiency in channel use is further demonstrated through statistical analysis.

It was argued that the optimal choice of the retransmission algorithm parameter is a function of the local vehicle density, specifically, the number of contending stations within range of the potential relay. In order to ascertain a measure of local density, this thesis presented a passive method to estimate the number of saturated stations on the network. This technique involves constantly listening to the channel, observing the number of idle slots between transmissions,

and subsequently applying Bayesian inference to estimate the number of contending stations on the network. A Markov model of the IEEE 802.11 DCF for broadcasts was presented, and by solving the model, the theoretical relationship between channel contention and inter-frame idle slot counts can be determined. This technique was compared to existing work, showing that the idle slot-based technique not only converges to steady state quicker, but also produces a smaller error in its estimate.

The effects of unsaturated stations on idle slot observations and collision probabilities are then explored. This thesis introduces a measure called "Equivalent Saturated Node" (ESN) to describe both a station's saturation level and the contention of a network. Through simulations, it was demonstrated that while the saturation level of a station does not form a linear relationship with either the idle slot counts or collision probabilities, the channel contention estimator is still capable of estimating the total channel saturation (in ESN) with only slight modification. It can be seen that the idle slot-based technique is more resilient to errors caused by unsaturated stations than the existing method. Statistical analysis conducted shows that the ESN measure is highly accurate in predicting packet non-reception.

Furthermore, an extended DCF model to account for unsaturated stations was presented in attempt to improve the performance of the estimator. However, this extended DCF model is found to be unviable, both due to inaccuracies in its predictions and extremely high complexity in solving the system of equations.

Based on the interference-aware geocast algorithm and the idle slot-based channel estimator, this thesis presented and evaluated a geocast system that can automatically adapt to channel contention. Channel estimates are fed into a load adaptation algorithm, which adjusts the retransmission parameter of the interference-aware geocast algorithm. Computer simulations show that this new, load-reactive geocast system is able to efficiently adapt to channel contention, and is able to self-correct suboptimal retransmission parameters calculated by

the load adaptation algorithm.

Finally, the discrepancies amongst the well known network simulators were outlined, with the impact of the bug in the ns-3 simulator evaluated. The simulators ns-2, ns-3 and OMNet++ all showed non-standard behaviour for IEEE 802.11 broadcast mode transmissions, with the error potentially skewing the outcomes of simulations. The non-standard behaviour in ns-3 is attributed to a longstanding reported bug. By comparing simulation data to theoretical predictions, it is demonstrated that this bug has a limited effect on the higher communication layers (above MAC layer). For simulations that rely on MAC layer statistics, a workaround that can be used to account for the misbehaviour was presented. This thesis also showed that similar misbehaviours still exist in ns-3.16 despite a code patch having been released that is supposed to have corrected the bug. Furthermore, the workaround proposed for the previous versions no longer applies to the misbehaviour in the new version.

9.1 Future work

The works presented in this thesis is merely a tip of the iceberg in the field of VANET communications. Potential extensions to the work in this thesis may include:

- **Investigate piggybacking/gossiping variants of retransmission protocol**. The retransmission metric presented in this thesis forms a useful metric to prioritise packets for retransmission. It is highly probable that the metric can be used for prioritising packets in the piggybacking protocol [125] or in setting the forwarding probabilities of gossiping protocols.

- **Investigate non-circular target areas**. The concept of taking into account areas being interfered with can be applied to non-circular areas such as rectangles in the case of road segments. When applied to non-circular areas, the effectiveness and behaviour of the algorithm may change. In ad-

dition the choice of parameter X will likely need to change. (Consider a long and narrow rectangle and a large circular transmission disc.) It may be possible that the parameter X may be insufficient for determining the desirability of a relay.

- **Derive analytical results for the geocast algorithms**. This thesis presented simulation results showing benefit of both the interference-aware algorithm and the adaptive geocast system. Analysis of these works can strengthen the claims in this thesis. Furthermore, graph-theoretic approximations of these algorithms that may arise from such investigations are likely to improve the scalability of the evaluation of similar algorithms.

- **Test the performance of the algorithms using vehicle layouts derived from realistic traffic simulators**. The works in this thesis investigated the performance of the algorithms using vehicle layouts that may not be very realistic due to the lack of software coupling ns-3 to realistic traffic simulators at that time. More realistic vehicle scenarios will allow stronger claims be made on the "real world" performance of these algorithms.

- **Alternative representation/analysis of unsaturated traffic**. This thesis represented unsaturated stations in a way that is difficult to apply practically. It would be useful to investigate the applicability of other existing (or develop alternative) methods of parameterising unsaturated stations in order to gain more insight into the DCF broadcast behaviour. This may also allow the load estimator to be more accurate at very low contention levels when the likelihood of extremely long inter-frame spaces may be present.

- **Investigate the validity of broadcast mode simulation in the most up-to-date versions of the simulators**. Despite the ns-3 simulator having been patched, misbehaviour in broadcast-mode transmissions is still present and the cause unknown. Other simulators are also continuously being improved. Given the importance of these simulators, a more in depth

understanding of these misbehaviours is needed both to correct the error and to accurately interpret simulation outcomes.

- **Investigate the cause of the self-correction behaviour in the load-reactive geocast algorithm**. The emergent behaviour observed had not yet been fully explained. An enhanced understanding of the cause can lead to this behaviour being exploited to improve the load adaptation algorithm.

- **Feasibility of multi-hop beacon forwarding**. Notwithstanding the methods presented in this thesis, it was observed that multi-hop beacon forwarding without explicit coordination (*e.g.* routing tree) is very expensive. It would be important to investigate the feasibility of such geocasting in the context of VANETs — Is it affordable? How would packet losses in single-hop beacons impact on safety applications?

Bibliography

[1] M. Lacage and T. R. Henderson, "Yet another network simulator," in Proceedings of *Proceeding from the 2006 Workshop on ns-2: The IP Network Simulator*. New York, NY, USA: ACM, 2006. [Online]. Available: http://doi.acm.org/10.1145/1190455.1190467

[2] D. Cosgrove *et al.*, "Estimating urban traffic and congestion cost trends for Australian cities," 2007.

[3] N. Grace *et al.*, "Transforming transportation through connectivity: ITS strategic research plan, 20102014 (progress update, 2012)," Technical Report, October 2012.

[4] Australian Transport Safety Bureau, "Road crash casualties and rates, Australia, 1925 to 2005," *ATSB, Canberra*, 2007. [Online]. Available: http://www.infrastructure.gov.au/roads/safety/publications/2008/pdf/1925_05_casualties.pdf

[5] Australian Transport Safety Bureau, "International road safety comparisons: the 2005 report," *ATSB Research and Analysis Report, Road Safety Monograph*, no. 19, 2007.

[6] J. G. Berry and J. E. Harrison, "Serious injury due to land transport accidents, Australia, 2005-06," *Injury Research and Statistics Series*, no. 42, July 2008.

[7] G. Weisbrod, D. Vary, and G. Treyz, "Measuring economic costs of urban traffic congestion to business," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1839, no. 2003, pp. 98–106.

[8] A. Sims and K. Dobinson, "The Sydney coordinated adaptive traffic (SCAT) system philosophy and benefits," *Vehicular Technology, IEEE Transactions on*, vol. 29, no. 2, pp. 130–137, 1980.

[9] P. Hunt *et al.*, "SCOOT — a traffic responsive method of coordinating signals," Technical Report, 1981.

[10] L. A. Klein, *Sensor Technologies and Data Requirements for ITS*. Artech House, 2001.

[11] T. Vaa, M. Penttinen, and I. Spyropoulou, "Intelligent transport systems and effects on road traffic accidents: state of the art," *Intelligent Transport Systems, IET*, vol. 1, no. 2, pp. 81–88, 2007.

[12] G. Goeldner, "Vision 2020," *NICTA intranet site*. [Online]. Available: http://wiki.inside.nicta.com.au/display/BARITIS/Vision+2020

[13] F. Kargl, "NICTA short course on security and privacy in intelligent transportation systems," July 2012.

[14] National Association of Emergency Medical Technicans (US), Prehospital Trauma Life Support Committee and American College of Surgeons, Committee on Trauma, *PHTLS: Prehospital Trauma Life Support*. Mosby JEMS/Elsevier, 2011.

[15] M. Paine *et al.*, "In-vehicle safety technologies: picking future winners!" in Proceedings of *Australasian Road Safety Research Policing Education Conference*, p. 17, November 2008.

[16] R. W. G. Anderson *et al.*, "Potential benefits of forward collision avoidance technology," Technical Report, April 2012.

[17] M. Jerbi, P. Marlier, and S. Senouci, "Experimental assessment of V2V and I2V communications," in Proceedings of *Mobile Adhoc and Sensor Systems. IEEE International Conference on*, pp. 1–6, October 2007.

[18] T. Yamamoto, "Activities on cooperative ITS in Japan," in Proceedings of *Fully Networked Car Conference*, 2011.

[19] CAMP Vehicle Safety Communications Consortium et al, "Vehicle safety communications project: Task 3 final report: Identify intelligent vehicle safety applications enabled by DSRC," *National Highway Traffic Safety Administration, US Department of Transportation, Washington DC*, 2005.

[20] E. Fitzgerald and B. Landfeldt, "A system for coupled road traffic utility maximisation and risk management using VANET," in Proceedings of *Intelligent Transportation Systems (ITSC), 15th International IEEE Conference on*, pp. 1880–1887. IEEE, 2012.

[21] C. Park and J. Lee, "Intelligent traffic control based on IEEE 802.11 DCF/PCF mechanisms at intersections," in Proceedings of *Vehicular Technology Conference (VTC Fall-2011), IEEE*, pp. 1–4, September 2011.

[22] Q. Xu *et al.*, "Vehicle-to-vehicle safety messaging in DSRC," in Proceedings of *Vehicular ad hoc Networks, 1st ACM International Workshop on.* ACM, 2004.

[23] T. ElBatt *et al.*, "Cooperative collision warning using dedicated short range wireless communications," in Proceedings of *Vehicular ad hoc Networks, 3rd International Workshop on*, pp. 1–9. New York, NY, USA: ACM, 2006. [Online]. Available: http://doi.acm.org/10.1145/1161064.1161066

[24] S. Biswas, R. Tatchikou, and F. Dion, "Vehicle-to-vehicle wireless communication protocols for enhancing highway traffic safety," *Communications Magazine, IEEE*, vol. 44, no. 1, pp. 74–82, January 2006.

[25] R. Sengupta *et al.*, "Cooperative collision warning systems: Concept definition and experimental implementation," *Intelligent Transportation Systems, Journal of*, vol. 11, no. 3, pp. 143–155, 2007.

[26] A. Tang and A. Yip, "Collision avoidance timing analysis of DSRC-based vehicles," *Accident Analysis & Prevention*, vol. 42, no. 1, pp. 182–195, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0001457509002024

[27] X. Chen, H. Refai, and X. Ma, "A quantitative approach to evaluate DSRC highway inter-vehicle safety communication," in Proceedings of *Global Telecommunications Conference, IEEE*, pp. 151–155, November 2007.

[28] S. Andrews and M. Cops, "Final report: Vehicle infrastructure integration proof of concept executive summary — Vehicle," Technical Report, May 2009.

[29] M. Sichitiu and M. Kihl, "Inter-vehicle communication systems: a survey," *Communications Surveys Tutorials, IEEE*, vol. 10, no. 2, pp. 88–105, quarter 2008.

[30] J. Yao *et al.*, "Improving cooperative positioning for vehicular networks," *Vehicular Technology, IEEE Transactions on*, vol. 60, no. 6, pp. 2810–2823, july 2011.

[31] M. Efatmaneshnik *et al.*, "A channel capacity perspective on cooperative positioning algorithms for VANET," in Proceedings of *Institute of Navigation — GNSS+*, pp. 1034–1041, 2009.

[32] "Intelligent transport systems (ITS): Communications architecture," *ETSI EN 302 665 V1.1.1 (2010-09)*, 2010.

[33] "Dedicated short range communications (DSRC) message set dictionary," *SAE J2735*, 2006.

[34] "IEEE draft guide for Wireless Access in Vehicular Environments (WAVE) — Architecture," *IEEE P1609.0/D5*, 2012.

[35] B. Villeforceix and S. Petti, "Communications in ITS for cooperative systems deployment," in Proceedings of *Fully Networked Car Conference*, 2011.

[36] "IEEE standard for Wireless Access in Vehicular Environments (WAVE) — Multi-channel operation," *IEEE Std 1609.4-2010 (Revision of IEEE Std 1609.4-2006)*, 2011.

[37] S. Eichler, "Performance evaluation of the IEEE 802.11p WAVE communication standard," in Proceedings of *Vehicular Technology Conference (VTC-2007 Fall). IEEE 66th*, pp. 2199–2203, Sep. 30–Oct. 3 2007.

[38] Z. Wang and M. Hassan, "How much of DSRC is available for non-safety use?" in Proceedings of *Vehicular Inter-networking, 5th ACM International Workshop on*, pp. 23–29. New York, NY, USA: ACM, 2008. [Online]. Available: http://doi.acm.org/10.1145/1410043.1410049

[39] K. Hong *et al.*, "Evaluation of multi-channel schemes for vehicular safety communications," in Proceedings of *Vehicular Technology Conference (VTC 2010-Spring), IEEE 71st*, pp. 1–5, May 2010.

[40] J. Misic, G. Badawy, and V. B. Misic, "Performance characterization for IEEE 802.11p network with single channel devices," *Vehicular Technology, IEEE Transactions on*, vol. 60, no. 4, pp. 1775–1787, May 2011.

[41] "IEEE standard for information technology — Local and metropolitan area networks — Specific requirements — Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications amendment 6: Wireless Access in Vehicular Environments," *IEEE Std 802.11p-2010 (Amendment to IEEE Std 802.11-2007 as amended by IEEE Std 802.11k-2008, IEEE Std 802.11r-2008, IEEE Std 802.11y-2008, IEEE Std 802.11n-2009, and IEEE Std 802.11w-2009)*, 2010.

[42] "Information technology — Telecommunications and information exchange between systems — Local and metropolitan area networks — Specific requirements part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications amendment 1: high-speed physical layer in the 5 GHz band," *ISO/IEC 8802-11:1999/Amd 1:2000(E); IEEE Std 802.11a-1999*, 2000.

[43] "IEEE standard for information technology — Telecommunications and information exchange between systems — Local and metropolitan area networks — Specific requirements part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications amendment 8: Medium Access Control (MAC) quality of service enhancements," *IEEE Std 802.11e-2005 (Amendment to IEEE Std 802.11, 1999 Edition (Reaff 2003))*, 2005.

[44] "IEEE standard for information technology — Telecommunications and information exchange between systems — Local and metropolitan area networks — Specific requirements part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications," *IEEE Std 802.11-2012 (Revision of IEEE Std 802.11-2007)*, 2012.

[45] G. Bianchi *et al.*, "Experimental assessment of the backoff behavior of commercial IEEE 802.11b network cards," in Proceedings of *Computer Communications, 26th IEEE International Conference on*, pp. 1181–1189, May 2007.

[46] G. Bianchi and I. Tinnirello, "Analysis of priority mechanisms based on differentiated inter frame spacing in CSMA-CA," in Proceedings of *Vehicular Technology Conference (VTC 2003-Fall), IEEE 58th*, vol. 3, pp. 1401–1405, October 2003.

[47] J. Gallardo *et al.*, "Analysis of the EDCA access mechanism for an IEEE 802.11e-compatible wireless LAN," in Proceedings of *Computers and Communications, IEEE Symposium on*, pp. 891–898, July 2008.

[48] J. Gallardo, D. Makrakis, and H. Mouftah, "Performance analysis of the EDCA medium access mechanism over the control channel of an IEEE 802.11p WAVE vehicular network," in Proceedings of *Communications, IEEE International Conference on*, pp. 1–6, June 2009.

[49] J. R. Gallardo, D. Makrakis, and H. T. Mouftah, "Mathematical analysis of EDCA's performance on the control channel of an IEEE 802.11p WAVE vehicular network," *Wireless Communication Networks, EURASIP Journal of*, vol. 2010, pp. 5:1–5:10, April 2010.

[50] R. Stanica, E. Chaput, and A.-L. Beylot, "Loss reasons in safety VANETs and implications on congestion control," in Proceedings of *Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks, 9th ACM Symposium on*, pp. 1–8. New York, NY, USA: ACM, 2012. [Online]. Available: http://doi.acm.org/10.1145/2387027.2387029

[51] S. Bastani, B. Landfeldt, and L. Libman, "On the reliability of safety message broadcasting urban vehicular ad hoc networks," in Proceedings of *Modelling, analysis and simulation of wireless and mobile systems, 14th ACM International Conference on*, pp. 307–316. New York, NY, USA: ACM, 2011. [Online]. Available: http://doi.acm.org/10.1145/2068897.2068951

[52] X. Ma and X. Chen, "Delay and broadcast reception rates of highway safety applications in vehicular ad hoc networks," in Proceedings of *Mobile Networking for Vehicular Environments*, pp. 85–90, May 2007.

[53] M. Torrent-Moreno, D. Jiang, and H. Hartenstein, "Broadcast reception rates and effects of priority access in 802.11-based vehicular ad-hoc

networks," in Proceedings of *Vehicular ad hoc networks, 1st ACM International Workshop on*, pp. 10–18. New York, NY, USA: ACM, 2004. [Online]. Available: http://doi.acm.org/10.1145/1023875.1023878

[54] F. Bai and H. Krishnan, "Reliability analysis of DSRC wireless communication for vehicle safety applications," in Proceedings of *Intelligent Transportation Systems Conference, IEEE*, pp. 355–362, September 2006.

[55] Y. Zhang and Y. Hwang, "Characterization of UHF radio propagation channels in tunnel environments for microcellular and personal communications," *Vehicular Technology, IEEE Transactions on*, vol. 47, no. 1, pp. 283–296, February 1998.

[56] J. Rustako, A.J. *et al.*, "Attenuation and diffraction effects from truck blockage of an 11-GHz line-of-sight microcellular mobile radio path," *Vehicular Technology, IEEE Transactions on*, vol. 40, no. 1, pp. 211–215, February 1991.

[57] R. Klingler, "Radio coverage for road and rail tunnels in the frequency range 75 to 1000 MHz," in Proceedings of *Vehicular Technology Conference, 41st IEEE*, pp. 433–438, May 1991.

[58] A. Paier *et al.*, "Overview of vehicle-to-vehicle radio channel measurements for collision avoidance applications," in Proceedings of *Vehicular Technology Conference (VTC 2010-Spring), IEEE 71st*, pp. 1–5, May 2010.

[59] R. Meireles *et al.*, "Experimental study on the impact of vehicular obstructions in VANETs," in Proceedings of *Vehicular Networking Conference, IEEE*, pp. 338–345, December 2010.

[60] M. Boban *et al.*, "Impact of vehicles as obstacles in vehicular ad hoc networks," *Selected Areas in Communications, IEEE Journal on*, vol. 29, no. 1, pp. 15–28, January 2011.

[61] T. Abbas *et al.*, "Measurement based shadow fading model for vehicle-to-vehicle network simulations," *submitted to IEEE Transactions on Vehicular Technology: Special Section on Vehicular Network and Communication System, From Laboratory into Reality*, vol. abs/1203.3370, 2014.

[62] L. Wischhof and H. Rohling, "Congestion control in vehicular ad hoc networks," in Proceedings of *Vehicular Electronics and Safety, IEEE International Conference on*, pp. 58–63, October 2005.

[63] M. M. Artimy, W. Robertson, and W. J. Phillips, "Assignment of dynamic transmission range based on estimation of vehicle density," in Proceedings of *Vehicular ad hoc Networks, 2nd ACM International Workshop on*, pp. 40–48. New York, NY, USA: ACM, 2005. [Online]. Available: http://doi.acm.org/10.1145/1080754.1080761

[64] M. Artimy, "Local density estimation and dynamic transmission-range assignment in vehicular ad hoc networks," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 8, no. 3, pp. 400–412, 2007.

[65] M. Torrent-Moreno *et al.*, "Vehicle-to-vehicle communication: Fair transmit power control for safety-critical information," *Vehicular Technology, IEEE Transactions on*, vol. 58, no. 7, pp. 3684–3703, September 2009.

[66] A. Sebastian *et al.*, "A multicast routing scheme for efficient safety message dissemination in VANET," in Proceedings of *Wireless Communications and Networking Conference, IEEE*, pp. 1–6, 18–21 April 2010.

[67] R. Stanica, E. Chaput, and A. Beylot, "Congestion control in CSMA-based vehicular networks: Do not forget the carrier sensing," in Proceedings of *Sensor, Mesh and Ad Hoc Communications and Networks, 9th Annual IEEE Communications Society Conference on*, pp. 650–658, June 2012.

[68] C. Chigan and J. Li, "A delay-bounded dynamic interactive power control algorithm for VANETs," in Proceedings of *Communications, IEEE International Conference on*, pp. 5849–5855, June 2007.

[69] Y. Xian and C.-T. Huang, "Traffic-aware geographic forwarding in vehicular ad hoc networks," in Proceedings of *Vehicular inter-networking, systems, and applications, Ninth ACM International Workshop on*, pp. 107–110. New York, NY, USA: ACM, 2012. [Online]. Available: http://doi.acm.org/10.1145/2307888.2307908

[70] C. Maihöfer, "A survey of geocast routing protocols," *Communications Surveys Tutorials, IEEE*, vol. 6, no. 2, pp. 32–42, Second 2004.

[71] F. Li and Y. Wang, "Routing in vehicular ad hoc networks: A survey," *Vehicular Technology Magazine, IEEE*, vol. 2, no. 2, pp. 12–22, June 2007.

[72] W. Chen *et al.*, "A survey and challenges in routing and data dissemination in vehicular ad hoc networks," *Wireless Communications and Mobile Computing*, vol. 11, no. 7, pp. 787–795, 2011. [Online]. Available: http://dx.doi.org/10.1002/wcm.862

[73] W. Peng and X.-C. Lu, "On the reduction of broadcast redundancy in mobile ad hoc networks," in Proceedings of *Mobile ad hoc networking & computing, 1st ACM International Symposium on*, pp. 129–130. Piscataway, NJ, USA: IEEE Press, 2000. [Online]. Available: http://dl.acm.org/citation.cfm?id=514151.514171

[74] F. Ros, P. Ruiz, and I. Stojmenovic, "Acknowledgment-based broadcast protocol for reliable and efficient data dissemination in vehicular ad hoc networks," *Mobile Computing, IEEE Transactions on*, vol. 11, no. 1, pp. 33–46, January 2012.

[75] D. Johnson, Y. Hu, and D. Maltz, "The dynamic source routing protocol (DSR) for mobile ad hoc networks for IPv4," RFC 4728

(Experimental), Internet Engineering Task Force, Feb. 2007. [Online]. Available: http://www.ietf.org/rfc/rfc4728.txt

[76] C. Perkins, E. Belding-Royer, and S. Das, "Ad hoc on-demand distance vector (AODV) routing," RFC 3561 (Experimental), Internet Engineering Task Force, Jul. 2003. [Online]. Available: http://www.ietf.org/rfc/rfc3561.txt

[77] C. Schwingenschlögl and T. Kosch, "Geocast enhancements of AODV for vehicular networks," *Mobile Computer Communication Review, SIGMOBILE*, vol. 6, no. 3, pp. 96–97, June 2002. [Online]. Available: http://doi.acm.org/10.1145/581291.581307

[78] T. Kosch, C. Schwingenschlogl, and L. Ai, "Information dissemination in multihop inter-vehicle networks," in Proceedings of *Intelligent Transportation Systems, IEEE 5th International Conference on*, pp. 685–690, 2002.

[79] S. Wang *et al.*, "A practical routing protocol for vehicle-formed mobile ad hoc networks on the roads," in Proceedings of *Intelligent Transportation Systems, IEEE Conference on*, pp. 161–166, September 2005.

[80] G. Liu *et al.*, "A routing strategy for metropolis vehicular communications," in Proceedings of *Information Networking. Networking Technologies for Broadband and Mobile Networks*, ser. Lecture Notes in Computer Science, H.-K. Kahng and S. Goto, Eds. Springer Berlin Heidelberg, 2004, vol. 3090, pp. 134–143. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-25978-7_14

[81] B.-C. Seet *et al.*, "A-STAR: A mobile ad hoc routing strategy for metropolis vehicular communications," in Proceedings of *NETWORKING 2004. Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications*, ser. Lecture Notes in Computer Science, N. Mitrou

et al., Eds. Springer Berlin Heidelberg, 2004, vol. 3042, pp. 989–999. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-24693-0_81

[82] S.-Y. Ni *et al.*, "The broadcast storm problem in a mobile ad hoc network," in Proceedings of *Mobile Computing and Networking, 5th Annual ACM/IEEE International Conference on*, pp. 151–162. New York, NY, USA: ACM, 1999. [Online]. Available: http://doi.acm.org/10.1145/313451.313525

[83] Y.-C. Tseng *et al.*, "The broadcast storm problem in a mobile ad hoc network," *Wireless Networks*, vol. 8, pp. 153–167, 2002. [Online]. Available: http://dx.doi.org/10.1023/A%3A1013763825347

[84] L. Briesemeister, L. Schäfers, and G. Hommel, "Disseminating messages among highly mobile hosts based on inter-vehicle communication," in Proceedings of *Intelligent Vehicles Symposium, IEEE*, pp. 522–527, 2000.

[85] W. Liao *et al.*, "Geogrid: A geocasting protocol for mobile ad hoc networks based on grid," *Internet Technology, Journal of*, vol. 1, no. 2, pp. 23–32, 2000.

[86] I. Stojmenovic, A. P. Ruhil, and D. K. Lobiyal, "Voronoi diagram and convex hull based geocasting and routing in wireless networks," *Wireless Communications and Mobile Computing*, vol. 6, no. 2, pp. 247–258, 2006. [Online]. Available: http://dx.doi.org/10.1002/wcm.384

[87] M. Mauve, J. Widmer, and H. Hartenstein, "A survey on position-based routing in mobile ad hoc networks," *Network, IEEE*, vol. 15, no. 6, pp. 30–39, November – December 2001.

[88] Y.-B. Ko and N. Vaidya, "Geocasting in mobile ad hoc networks: location-based multicast algorithms," in Proceedings of *Mobile Computing Systems and Applications, Second IEEE Workshop on*, pp. 101–110, February 1999.

[89] M. Boban *et al.*, "Exploiting the height of vehicles in vehicular communication," in Proceedings of *Vehicular Networking Conference, IEEE*, pp. 163–170, November 2011.

[90] Z. Haas, J. Halpern, and L. Li, "Gossip-based ad hoc routing," in Proceedings of *Computer and Communications Societies, 21st Annual Joint Conference of the IEEE*, vol. 3, pp. 1707–1716, 2002.

[91] Z. J. Haas, J. Y. Halpern, and L. Li, "Gossip-based ad hoc routing," *Networking, IEEE/ACM Transactions on*, vol. 14, no. 3, pp. 479–491, June 2006. [Online]. Available: http://dx.doi.org/10.1109/TNET.2006.876186

[92] R. Chandra, V. Ramasubramanian, and K. Birman, "Anonymous Gossip: improving multicast reliability in mobile ad-hoc networks," in Proceedings of *Distributed Computing Systems, 21st International Conference on*, pp. 275–283, April 2001.

[93] J. Luo, P. Eugster, and J.-P. Hubaux, "Route driven gossip: probabilistic reliable multicast in ad hoc networks," in Proceedings of *Computer and Communications Societies, 22nd Annual Joint Conference of the IEEE*, vol. 3, pp. 2229–2239, March–April 2003.

[94] K. P. Birman *et al.*, "Bimodal multicast," *Computer Systems, ACM Transactions on*, vol. 17, no. 2, pp. 41–88, May 1999. [Online]. Available: http://doi.acm.org/10.1145/312203.312207

[95] P. Kyasanur, R. R. Choudhury, and I. Gupta, "Smart gossip: An adaptive gossip-based broadcasting service for sensor networks," in Proceedings of *Mobile Adhoc and Sensor Systems, IEEE International Conference on*, pp. 91–100, October 2006.

[96] B. Bako *et al.*, "Advanced adaptive gossiping using 2-hop neighborhood information," in Proceedings of *Global Telecommunications Conference, IEEE*, pp. 1–6, Nov. 30–Dec. 4 2008.

[97] B. Bako *et al.*, "Adaptive topology based gossiping in VANETs using position information," in Proceedings of *Mobile Ad-Hoc and Sensor Networks*, ser. Lecture Notes in Computer Science, H. Zhang *et al.*, Eds. Springer Berlin Heidelberg, 2007, vol. 4864, pp. 66–78. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-77024-4_8

[98] B. Bako *et al.*, "Evaluation of position based gossiping for VANETs in an intersection scenario," in Proceedings of *Networked Computing and Advanced Information Management, 4th International Conference on*, vol. 1, pp. 397–402, September 2008.

[99] Y. Mylonas, M. Lestas, and A. Pitsillides, "Speed adaptive probabilistic flooding in cooperative emergency warning," in Proceedings of *Wireless Internet, International Conference on*, pp. 1–7, ICST, Brussels, Belgium, 2008.

[100] B. Bako *et al.*, "Optimized position based gossiping in VANETs," in Proceedings of *Vehicular Technology Conference (VTC 2008-Fall). IEEE 68th*, pp. 1–5, September 2008.

[101] L. Briesemeister and G. Hommel, "Role-based multicast in highly mobile but sparsely connected ad hoc networks," in Proceedings of *Mobile ad hoc networking & computing, 1st ACM International Symposium on*, pp. 45–50. Piscataway, NJ, USA: IEEE Press, 2000. [Online]. Available: http://dl.acm.org/citation.cfm?id=514151.514159

[102] N. Wisitpongphan *et al.*, "Broadcast storm mitigation techniques in vehicular ad hoc networks," *Wireless Communications, IEEE*, vol. 14, no. 6, pp. 84–94, December 2007.

[103] B. Karp and H. T. Kung, "GPSR: greedy perimeter stateless routing for wireless networks," in Proceedings of *Mobile computing and networking, 6th Annual International Conference on*, pp. 243–

254. New York, NY, USA: ACM, 2000. [Online]. Available: http://doi.acm.org/10.1145/345910.345953

[104] C. Lochert *et al.*, "Geographic routing in city scenarios," *Mobile Computer Communication Review, ACM SIGMOBILE*, vol. 9, no. 1, pp. 69–72, January 2005. [Online]. Available: http://doi.acm.org/10.1145/1055959.1055970

[105] K. Lee *et al.*, "Enhanced perimeter routing for geographic forwarding protocols in urban vehicular scenarios," in Proceedings of *GLOBECOM Workshops, 2007 IEEE*, pp. 1–10, November 2007.

[106] E. Kranakis, H. Singh, and J. Urrutia, "Compass routing on geometric networks," in Proceedings of *Proc. 11 th Canadian Conference on Computational Geometry*, 1999.

[107] P. Bose *et al.*, "Routing with guaranteed delivery in ad hoc wireless networks," *Wireless Networks*, vol. 7, pp. 609–616, 2001. [Online]. Available: http://dx.doi.org/10.1023/A%3A1012319418150

[108] F. Kuhn, R. Wattenhofer, and A. Zollinger, "Worst-case optimal and average-case efficient geometric ad-hoc routing," in Proceedings of *Mobile ad hoc networking & computing, 4th ACM International Symposium on*, pp. 267–278. New York, NY, USA: ACM, 2003. [Online]. Available: http://doi.acm.org/10.1145/778415.778447

[109] B. Leong, S. Mitra, and B. Liskov, "Path vector face routing: geographic routing with local face information," in Proceedings of *Network Protocols, 13th IEEE International Conference on*, November 2005.

[110] A. Odorizzi and G. Mazzini, "M-GeRaF: A reliable random forwarding geographic routing protocol in multisink ad hoc and sensor networks," in Proceedings of *Intelligent Signal Processing and Communication Systems, International Symposium on*, pp. 416–419, November 2007.

[111] M. Zorzi and R. Rao, "Geographic Random Forwarding (GeRaF) for ad hoc and sensor networks: multihop performance," *Mobile Computing, IEEE Transactions on*, vol. 2, no. 4, pp. 337–348, October – December 2003.

[112] G. Korkmaz *et al.*, "Urban Multi-hop Broadcast protocol for inter-vehicle communication systems," in Proceedings of *Vehicular ad hoc networks, 1st ACM International Workshop on*, pp. 76–85. New York, NY, USA: ACM, 2004. [Online]. Available: http://doi.acm.org/10.1145/1023875.1023887

[113] G. Korkmaz, E. Ekici, and F. Özgüner, "An efficient fully ad-hoc multi-hop broadcast protocol for inter-vehicular communication systems," in Proceedings of *Communications, IEEE International Conference on*, vol. 1, pp. 423–428, June 2006.

[114] M. Taha and Y. Hasan, "VANET-DSRC protocol for reliable broadcasting of life safety messages," in Proceedings of *Signal Processing and Information Technology, IEEE International Symposium on*, pp. 104–109, December 2007.

[115] E. Fasolo, A. Zanella, and M. Zorzi, "An effective broadcast scheme for alert message propagation in vehicular ad hoc networks," in Proceedings of *Communications, IEEE International Conference on*, vol. 9, June 2006.

[116] M. Barradi, A. Hafid, and S. Aljahdali, "Highway multihop broadcast protocols for vehicular networks," in Proceedings of *Communications, IEEE International Conference on*, pp. 5296–5300, June 2012.

[117] M. Máté and R. Vida, "Probability-based information dissemination in urban environments," in Proceedings of *the Proceedings of EUNICE*, pp. 51–58, 2008.

[118] M. Maté and R. Vida, "Reliable gossiping in urban environments," in Proceedings of *Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd*, pp. 1–5, September 2010.

[119] Y. Sung and M. Lee, "Light-weight reliable broadcast message delivery for vehicular ad-hoc networks," in Proceedings of *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th*, pp. 1–6, May 2012.

[120] O. Tonguz, N. Wisitpongphan, and F. Bai, "DV-CAST: A distributed vehicular broadcast protocol for vehicular ad hoc networks," *Wireless Communications, IEEE*, vol. 17, no. 2, pp. 47–57, April 2010.

[121] R. Jain, A. Puri, and R. SenGupta, "Geographical routing using partial information for wireless ad hoc networks," *Personal Communications, IEEE*, vol. 8, no. 1, pp. 48–57, February 2001.

[122] A. T. Hoang and M. Motani, "Collaborative broadcasting and compression in cluster-based wireless sensor networks," *Sensor Networks, ACM Transactions on*, vol. 3, no. 3, August 2007. [Online]. Available: http://doi.acm.org/10.1145/1267060.1267065

[123] R. Santos *et al.*, "Performance evaluation of routing protocols in vehicular ad-hoc networks," *Ad Hoc and Ubiquitous Computing, International Journal of*, vol. 1, no. 1, pp. 80–91, 2005. [Online]. Available: http://inderscience.metapress.com/content/BR9WGECYX6JGWD0X

[124] D. Jiang *et al.*, "Design of 5.9 GHz DSRC-based vehicular safety communication," *Wireless Communications, IEEE*, vol. 13, no. 5, pp. 36–43, October 2006.

[125] L. Yang, J. Guo, and Y. Wu, "Piggyback cooperative repetition for reliable broadcasting of safety messages in VANETs," in Proceedings of *Consumer Communications and Networking Conference, 6th IEEE*, pp. 1–5, January 2009.

[126] L. Wischoff *et al.*, "SOTIS — a self-organizing traffic information system," in Proceedings of *Vehicular Technology Conference (VTC 2003-Spring), 57th IEEE Semiannual*, vol. 4, pp. 2442–2446, April 2003.

[127] B. Scheuermann *et al.*, "A fundamental scalability criterion for data aggregation in VANETs," in Proceedings of *Mobile computing and networking, 15th Annual International Conference on*, pp. 285–296. New York, NY, USA: ACM, 2009. [Online]. Available: http://doi.acm.org/10.1145/1614320.1614352

[128] S. Dietzel *et al.*, "A fuzzy logic based approach for structure-free aggregation in vehicular ad-hoc networks," in Proceedings of *Vehicular Inter-networking, 6th ACM International Workshop on*, pp. 79–88. New York, NY, USA: ACM, 2009. [Online]. Available: http://doi.acm.org/10.1145/1614269.1614283

[129] M. Caliskan, D. Graupner, and M. Mauve, "Decentralized discovery of free parking places," in Proceedings of *Vehicular ad hoc networks, 3rd International Workshop on*, pp. 30–39. New York, NY, USA: ACM, 2006. [Online]. Available: http://doi.acm.org/10.1145/1161064.1161070

[130] C. Lochert, B. Scheuermann, and M. Mauve, "Probabilistic aggregation for data dissemination in VANETs," in Proceedings of *Vehicular ad hoc networks, 4th ACM International Workshop on*, pp. 1–8. New York, NY, USA: ACM, 2007. [Online]. Available: http://doi.acm.org/10.1145/1287748.1287750

[131] Q. Yu and D. Liu, "Disseminate warning message in VANETs based on predicting the interval of vehicles," in Proceedings of *Frontier of Computer Science and Technology, Fifth International Conference on*, pp. 559–564, August 2010.

[132] M. Bilal, P. Chan, and P. Pillai, "A fastest multi-hop routing scheme for information dissemination in vehicular communication systems," in Proceedings of *Software, Telecommunications and Computer Networks International Conference on*, pp. 35–41, September 2010.

[133] J. LeBrun *et al.*, "Knowledge-based opportunistic forwarding in vehicular wireless ad hoc networks," in Proceedings of *Vehicular Technology Conference (VTC 2005-Spring), IEEE 61st*, vol. 4, pp. 2289–2293, May–1 June 2005.

[134] R. Reinders *et al.*, "Contention window analysis for beaconing in VANETs," in Proceedings of *Wireless Communications and Mobile Computing Conference, 7th International*, pp. 1481–1487, July 2011.

[135] Y. Tay, K. Jamieson, and H. Balakrishnan, "Collision-minimizing CSMA and its applications to wireless sensor networks," *Selected Areas in Communications, IEEE Journal on*, vol. 22, no. 6, pp. 1048–1057, August 2004.

[136] G. Bianchi and I. Tinnirello, "Kalman filter estimation of the number of competing terminals in an IEEE 802.11 network," in Proceedings of *Computer and Communications Societies, 22nd Annual Joint Conference of the IEEE*, vol. 2, pp. 844–852, March-3 April 2003.

[137] A. Toledo, T. Vercauteren, and X. Wang, "Adaptive optimization of IEEE 802.11 DCF based on Bayesian estimation of the number of competing terminals," *Mobile Computing, IEEE Transactions on*, vol. 5, no. 9, pp. 1283–1296, September 2006.

[138] T. Vercauteren, A. L. Toledo, and X. Wang, "Batch and sequential Bayesian estimators of the number of active terminals in an IEEE 802.11 network," *Signal Processing, IEEE Transactions on*, vol. 55, no. 2, pp. 437–450, February 2007.

[139] J.-S. Kim, E. Serpedin, and D.-R. Shin, "Improved particle filtering-based estimation of the number of competing stations in IEEE 802.11 networks," *Signal Processing Letters, IEEE*, vol. 15, pp. 87–90, 2008.

[140] L. Bononi, M. Conti, and L. Donatiello, "Design and performance evaluation of a Distributed Contention Control (DCC) mechanism

for IEEE 802.11 wireless local area networks," in Proceedings of *Wireless mobile multimedia, 1st ACM International Workshop on*, pp. 59–67. New York, NY, USA: ACM, 1998. [Online]. Available: http://doi.acm.org/10.1145/288338.288378

[141] M. Heusse *et al.*, "Idle sense: an optimal access method for high throughput and fairness in rate diverse wireless LANs," *Computer Communication Review, ACM SIGCOMM*, vol. 35, pp. 121–132, August 2005.

[142] A. Richter, R. Thomae, and T. Taga, "Directional measurement and analysis of propagation path variations in a street micro-cell scenario," in Proceedings of *Vehicular Technology Conference (VTC 2003-Spring), 57th IEEE Semiannual*, vol. 1, pp. 246–250, April 2003.

[143] J. Yin *et al.*, "Performance evaluation of safety applications over DSRC vehicular ad hoc networks," in Proceedings of *Vehicular ad hoc Networks, 1st ACM International Workshop on*, pp. 1–9. ACM, 2004.

[144] J. H. Kim and J. K. Lee, "Capture effects of wireless CSMA/CA protocols in rayleigh and shadow fading channels," *Vehicular Technology, IEEE Transactions on*, vol. 48, no. 4, pp. 1277–1286, July 1999.

[145] K. Ramachandran *et al.*, "Experimental analysis of broadcast reliability in dense vehicular networks," in Proceedings of *Vehicular Technology Conference (VTC-2007 Fall), IEEE 66th*, pp. 2091–2095, Sept. 30–Oct. 3 2007.

[146] L. Bononi, M. Conti, and E. Gregori, "Runtime optimization of IEEE 802.11 wireless LANs performance," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 15, no. 1, pp. 66–80, January 2004.

[147] F. Calì, M. Conti, and E. Gregori, "Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit," *Networking, IEEE/ACM Transactions on*, vol. 8, no. 6, pp. 785–799, December 2000.

[148] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *Selected Areas in Communications, IEEE Journal on*, vol. 18, no. 3, pp. 535–547, March 2000.

[149] G. Bianchi, L. Fratta, and M. Oliveri, "Performance evaluation and enhancement of the CSMA/CA MAC protocol for 802.11 wireless LANs," in Proceedings of *Personal, Indoor and Mobile Radio Communications, 7th IEEE International Symposium on*, vol. 2, pp. 392–396, October 1996.

[150] F. Daneshgaran *et al.*, "Unsaturated throughput analysis of IEEE 802.11 in presence of non ideal transmission channel and capture effects," *Wireless Communications, IEEE Transactions on*, vol. 7, no. 4, pp. 1276–1286, April 2008.

[151] K. Duffy, D. Malone, and D. Leith, "Modeling the 802.11 distributed coordination function in non-saturated conditions," *Communications Letters, IEEE*, vol. 9, no. 8, pp. 715–717, August 2005.

[152] D. Malone, K. Duffy, and D. Leith, "Modeling the 802.11 distributed coordination function in nonsaturated heterogeneous conditions," *Networking, IEEE/ACM Transactions on*, vol. 15, no. 1, pp. 159–172, February 2007.

[153] V. Vishnevsky and A. Lyakhov, "802.11 LANs: Saturation throughput in the presence of noise," *NETWORKING 2002: Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications*, vol. 2345, pp. 1008–1019, 2006.

[154] D. Malone, I. Dangerfield, and D. Leith, "Verification of common 802.11 MAC model assumptions," in Proceedings of *Passive and Active Network Measurement, Internal Conference on*, ser. Lecture Notes in Computer Science, S. Uhlig, K. Papagiannaki, and O. Bonaventure, Eds. Springer Berlin / Heidelberg, 2007, vol. 4427, pp. 63–72.

[155] K. Huang *et al.*, "Investigating the validity of IEEE 802.11 MAC modeling hypotheses," in Proceedings of *Personal, Indoor and Mobile Radio Communications, IEEE 19th International Symposium on*, pp. 1–6, September 2008.

[156] K. Huang, K. Duffy, and D. Malone, "On the validity of IEEE 802.11 MAC modeling hypotheses," *Networking, IEEE/ACM Transactions on*, vol. 18, no. 6, pp. 1935–1948, December 2010.

[157] X. Ma, X. Chen, and H. Refai, "Unsaturated performance of IEEE 802.11 broadcast service in vehicle-to-vehicle networks," in Proceedings of *Vehicular Technology Conference (VTC-2007 Fall), IEEE 66th*, pp. 1957–1961, September – October 2007.

[158] M. van Eenennaam, A. Remke, and G. Heijenk, "An analytical model for beaconing in VANETs," in Proceedings of *Vehicular Networking Conference (VNC), 2012 IEEE*, pp. 9–16, November 2012.

[159] Q. Pang, S. Liew, and V. Leung, "Design of an effective loss-distinguishable MAC protocol for 802.11 WLAN," *Communications Letters, IEEE*, vol. 9, no. 9, pp. 781–783, September 2005.

[160] Q. Pang, V. Leung, and S. Liew, "A rate adaptation algorithm for IEEE 802.11 WLANs based on MAC-layer loss differentiation," in Proceedings of *Broadband Networks, 2nd International Conference on*, vol. 1, pp. 659–667, October 2005.

[161] H. Ma, S. Y. Shin, and S. Roy, "Optimizing throughput with carrier sensing adaptation for IEEE 802.11 mesh networks based on loss differentiation," in Proceedings of *Communications, IEEE International Conference on*, pp. 4967–4972, June 2007.

[162] D. Malone, P. Clifford, and D. Leith, "MAC layer channel quality measurement in 802.11," *Communications Letters, IEEE*, vol. 11, no. 2, pp. 143–145, February 2007.

[163] D. Qiao and S. Choi, "Goodput enhancement of IEEE 802.11a wireless LAN via link adaptation," in Proceedings of *Communications, IEEE International Conference on*, vol. 7, pp. 1995–2000, 2001.

[164] D. Qiao, S. Choi, and K. Shin, "Goodput analysis and link adaptation for IEEE 802.11a wireless LANs," *Mobile Computing, IEEE Transactions on*, vol. 1, no. 4, pp. 278–292, Oct–Dec 2002.

[165] N. Samaraweera, "Non-congestion packet loss detection for TCP error recovery using wireless links," *Communications, IEE Proceedings*, vol. 146, no. 4, pp. 222–230, August 1999.

[166] Y. Tobe *et al.*, "Achieving moderate fairness for UDP flows by path-status classification," in Proceedings of *Local Computer Networks, 25th Annual IEEE Conference on)*, pp. 252–261. Washington, DC, USA: IEEE Computer Society, 2000. [Online]. Available: http://dl.acm.org/citation.cfm?id=788015.788541

[167] S. Cen, P. C. Cosman, and G. M. Voelker, "End-to-end differentiation of congestion and wireless losses," *Networking, IEEE/ACM Transactions on*, vol. 11, no. 5, pp. 703–717, October 2003. [Online]. Available: http://dx.doi.org/10.1109/TNET.2003.818187

[168] S. Kurkowski, T. Camp, and M. Colagrosso, "MANET simulation studies: the incredibles," *Mobile Computer Communication Review, ACM SIGMOBILE*, vol. 9, no. 4, pp. 50–61, October 2005. [Online]. Available: http://doi.acm.org/10.1145/1096166.1096174

[169] D. Johnson, "Validation of wireless and mobile network models and simulation," in Proceedings of *DARPA/NIST Network Simulation Validation Workshop*, 1999.

[170] J. Liu *et al.*, "Simulation validation using direct execution of wireless ad-hoc routing protocols," in Proceedings of *Parallel and Distributed Simulation, 18th Workshop on*, pp. 7–16, May 2004.

[171] A. Rachedi *et al.*, "Wireless network simulators relevance compared to a real testbed in outdoor and indoor environments," *International Journal of Autonomous and Adaptive Communications Systems*, vol. 5, pp. 88–101, 2012. [Online]. Available: http://inderscience.metapress.com/content/G64Q42888G038R33

[172] G. Pei and T. Henderson, "Validation of ns-3 802.11b PHY model," *Online: http://www.nsnam.org/~pei/80211b.pdf*, 2009.

[173] S. Ivanov, A. Herms, and G. Lukas, "Experimental validation of the ns-2 wireless model using simulation, emulation, and real network," *Communication in Distributed Systems (KiVS), ITG-GI Conference*, pp. 1–12, February 26 – March 2 2007.

[174] U. M. Colesanti, C. Crociani, and A. Vitaletti, "On the accuracy of OMNeT++ in the wireless sensor networks domain: simulation vs. testbed," in Proceedings of *Performance evaluation of wireless ad hoc, sensor,and ubiquitous networks, 4th ACM Workshop on*, pp. 25–31. New York, NY, USA: ACM, 2007. [Online]. Available: http://doi.acm.org/10.1145/1298197.1298203

[175] D. Cavin, Y. Sasson, and A. Schiper, "On the accuracy of MANET simulators," in Proceedings of *Principles of mobile computing, 2nd ACM International Workshop on*, pp. 38–43. New York, NY, USA: ACM, 2002. [Online]. Available: http://doi.acm.org/10.1145/584490.584499

[176] M. Bredel and M. Bergner, "On the accuracy of IEEE 802.11g wireless LAN simulations using OMNeT++," in Proceedings of *Simulation Tools and Techniques, 2nd International Conference on*, pp. 81:1–81:5. ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009. [Online]. Available: http://dx.doi.org/10.4108/ICST.SIMUTOOLS2009.5585

[177] N. Baldo *et al.*, "Validation of the IEEE 802.11 MAC model in the ns3 simulator using the EXTREME testbed," in Proceedings of *Simulation Tools and Techniques, 3rd International Conference on*, pp. 64:1–64:9. ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2010. [Online]. Available: http://dx.doi.org/10.4108/ICST.SIMUTOOLS2010.8705

[178] E. W. Weisstein, "Circle-circle intersection," *MathWorld — A Wolfram Web Resource*. [Online]. Available: http://mathworld.wolfram.com/Circle-CircleIntersection.html

[179] M. Torrent-Moreno *et al.*, "IEEE 802.11-based one-hop broadcast communications: understanding transmission success and failure under different radio propagation environments," in Proceedings of *Proceedings of the 9th ACM international symposium on Modeling analysis and simulation of wireless and mobile systems*, pp. 68–77. ACM, 2006.

[180] H. T. Friis, "A note on a simple transmission formula," *Proceedings of the Institute of Radio Engineers*, vol. 34, no. 5, pp. 254–256, 1946.

[181] J. Gettys and K. Nichols, "Bufferbloat: dark buffers in the internet," *Communications, ACM*, vol. 55, no. 1, pp. 57–65, January 2012. [Online]. Available: http://doi.acm.org/10.1145/2063176.2063196

[182] B. Buchberger, "Theoretical basis for the reduction of polynomials to canonical forms," *Symbolic and Algebraic Manipulation Bulletin, ACM Special Interest Group in*, vol. 39, pp. 19–24, August 1976.

[183] E. Mayr, "Some complexity results for polynomial ideals," *Complexity, Journal of*, vol. 13, no. 3, pp. 301–384, 1997.

[184] P. L'ecuyer, "Good parameters and implementations for combined multiple recursive random number generators," *Operations Research*, vol. 47, no. 1, pp. 159–164, 1999.

[185] M. Torrent-Moreno, M. Killat, and H. Hartenstein, "The challenges of robust inter-vehicle communications," in Proceedings of *Vehicular Technology Conference (VTC-2005 Fall), IEEE 62nd*, vol. 1, pp. 319–323, September 2005.

[186] T. Bingmann, "Bug 555 — DCF immediate access bug." [Online]. Available: https://www.nsnam.org/bugzilla/show_bug.cgi?id=555