ACHIEVING DYNAMIC ROAD TRAFFIC MANAGEMENT BY DISTRIBUTED RISK ESTIMATION IN VEHICULAR NETWORKS

THE UNIVERSITY OF
SYDNEY

A thesis submitted in fulfilment of the requirements for the
degree of Doctor of Philosophy in the Faculty of Engineering and Information
Technologies at
The University of Sydney

Emma Fitzgerald

August 2013

Abstract

In this thesis I develop a model for a dynamic and fine-grained approach to traffic management based around the concept of a risk limit: an acceptable or allowable level of accident risk which vehicles must not exceed. Using a vehicular network to exchange risk data, vehicles calculate their current level of accident risk and determine their behaviour in a distributed fashion in order to meet this limit. I conduct experimental investigations to determine the effectiveness of this model, showing that it is possible to achieve gains in road system utility in terms of average vehicle speed and overall throughput whilst maintaining the accident rate. I also extend this model to include risk-aware link choice and social link choice, in which vehicles make routing decisions based on both their own utility and the utility of following vehicles.

I develop a coupled risk estimation algorithm in which vehicles use not only their own risk calculations but also estimates received from neighbouring vehicles in order to arrive at a final risk value. I then analyse the performance of this algorithm in terms of its convergence rate and bandwidth usage and examine how to manage the particular characteristics of a vehicular ad-hoc network, such as its dynamic topology and high node mobility. I then implement a variable-rate beaconing scheme to provide a trade-off between risk estimate error and network resource usage.

Acknowledgements

I would like to thank Björn, for his guidance, advice and inspiration; Alan, for stepping in to help at a late hour; Andrew, for suffering my stress with good humour and love; my family, for always believing in me; Kaz, for being my canary in the mine; Nicky and Georgina, for emergency chocolate and companionship on the journey; and my friends, for always being there to provide a few hours of much-needed company and escapism.

Contents

List of Tables

List of Figures

Part I

Introduction

Chapter 1

Introduction

This thesis will present a new approach to traffic management that utilises emerging technologies such as vehicular networks to yield a fine-grained, dynamic means of balancing the safety and utility of the road system, with the aim of improving road system utility whilst maintaining the accident rate. With computational power in each vehicle and communications between vehicles comes the availability of individualised, up-to-date information about the current situation in which each vehicle finds itself and the ability to process and act on that information.

A core problem in traffic management is to balance accident risk with the utility of the road system, which consists of individual vehicle speeds and overall traffic throughput in the network. To achieve this, we introduce the concept of a risk limit, analogous to a speed limit. The risk limit represents an acceptable level of accident risk which all vehicles must maintain. Below this level, vehicles may seek to optimise utility as much as possible so long as they do not exceed the risk limit.

We develop a new model for traffic management that uses the concept of risk limits. Each vehicle continuously calculates its current risk level based on the risk factors in effect at the time and compares this value to the risk limit. If the current level of risk is higher than the acceptable limit, the vehicle must act in a way that reduces its risk. If the level of risk is currently below the threshold, however, the vehicle may instead employ behaviours that increase its utility.

1.1 Aims

The purpose of the work presented in this thesis is to harness vehicular ad-hoc networks (VANETs) to improve traffic management. The aim here is to maintain a sufficiently low level of accident risk while increasing road system utility. Here, utility pertains to both the individual level — as represented by measures such as trip time or vehicle speed — and the larger, system level, which is concerned with overall traffic throughput.

By taking into account the most relevant and specific information possible about the current situation and about individual vehicles and drivers, we aim to provide a more flexible system that has utility gains over existing methods of risk mitigation without increasing the traffic accident rate. This flexibility can permit a wider range of responses to risk and can allow these responses to be tailored to specific situations. An important effect of this strategy is the potential to allow drivers with impairments — who are currently considered to be too high a risk to qualify for drivers' licenses — to be able to drive, as all of the factors affecting them can be taken into account and mitigated.

A further aim is to understand the effects of using such a system and to investigate its feasibility and reliability. This includes measuring the effects of this system on safety and utility as well as investigating how to best use the network to distribute risk information and how each vehicle and driver should determine appropriate behaviour to take in any given situation. Moreover, it is important to establish how the system can be made robust to network phenomena such as contention, varying signal strength, failed transmissions and a rapidly changing network topology such as is typical in vehicular networks. We also examine how such a system might be implemented and what the requirements on the network are for it to be effective.

1.2 Contributions

Risk limit model for traffic management

The main contribution of this thesis is the risk limit model for traffic accident risk management presented in Chapter 4, which enables road system utility to be increased whilst maintaining the accident rate at an acceptable level. This model centres around the concept of a risk limit as the primary mechanism for managing accident risk and balancing it with utility requirements for the road system. This is a new approach to traffic management and contrasts with existing methods of risk mitigation, which are for the most part static and coarse-grained, applying in all circumstances to all vehicles rather than being tailored to an individual situation, driver or vehicle.

Method for determining current, individualised risk levels

In order to use a risk limit effectively, it is necessary to have a means of estimating the current risk level. This thesis examines how this can be achieved and present a method for doing so, utilising real-world data collected from traffic authorities in NSW, Australia. This relies on a way of formulating accident risk I have developed that is generic to any situation and set of risk factors and which can be extended to incorporate more risk factors as more and better accident data becomes available.

Model for modifying vehicle behaviour to mitigate risk and increase utility

In order to use a risk limit for traffic management, vehicles must have a means of meeting the limit. This requires a means of modifying vehicle behaviour based on how their current risk level compares with the risk limit. I present a model for vehicle behaviour informed by the risk limit and current estimation. This model is again generic and modular, able to be expanded as new behaviours are modelled and added to the system. Additionally, I investigate some possible behaviours — speed, headway, lane choice and link choice — by conducting simulation-based experiments. The aim of these experiments was to determine the behaviours' effects on accident risk in order to incorporate them in the risk limit system and utilise them for improving utility and controlling accident risk.

Risk-aware and socially-aware link choice algorithm and experiments

I develop an algorithm for risk-aware link choice and examine the effects of this on traffic throughput and vehicle speeds. The link choice model is then extended to include social awareness, such that vehicles will consider not only their own utility but also the effects their choices have on other vehicles' utility. I have measured the effect of varying the level of social awareness in order to find the optimal level of self-interest for vehicles in terms of overall throughput and vehicle speeds.

Experimental investigations into the effectiveness of the risk limit system for traffic management

I conducted experimental investigations to determine the effectiveness of the risk limit system in managing traffic, that is, controlling accident risk and maximising utility. This involved the combination of risk estimation based on an understanding and measurement of pertinent risk factors, a risk limit representing a desirable or acceptable level of accident risk, and a means for adapting vehicle behaviour to their current risk level in order to meet this limit. Using this system, I show that it is possible to achieve gains in the utility of the road system while maintaining the accident rate at or below current levels.

Coupled risk estimation algorithm: convergence proof and experimental investigations

The system was further expanded to include an algorithm for coupled risk estimation, in which vehicles exchange information about accident risk and modify their risk estimates based on information received from other vehicles. I prove the convergence of this algorithm and have conducted experimental investigations into demonstrating its effectiveness in controlling accident risk in the presence of hazards localised to a particular section of the road, this being a situation that particularly calls for such an approach. I have also investigated the information propagation properties of this algorithm to ensure they match the theoretical design and have measured convergence rates and bandwidth requirements, showing them to be feasible for implementation and use.

Analysis of the coupled risk estimation algorithm under different network conditions and variable beaconing rate scheme

I also conducted a more in-depth analysis into the networking behaviour of the coupled risk estimation algorithm. I undertook experimental investigations into how the convergence rate was affected by network size and node density, and how the beaconing rate affected error levels in the algorithm's outputs. I then developed a variable beaconing rate scheme in order to find a balance between a fast convergence rate — yielding low error levels, even when the inputs to the algorithm changed rapidly and unpredictably — and network resource usage.

Literature survey

In Chapter 3, I conduct a literature survey of the research areas that influence the development and implementation of the risk limit system. This requires input from a wide range of fields including human factors, road safety, driver and vehicle modelling, advanced driver assistance systems, co-operative driving, and vehicular ad-hoc networks. I synthesise results from these areas to give an overview of the literature and how it affects the systems and models presented in the rest of this thesis. This survey has bearing on future work done in the domain of dynamic, individual-based traffic management.

1.3 Outline

The rest of this thesis is organised as follows. Chapter 2 gives an overview of the motivation for this research, from the perspectives of road safety and road system utility, in particular in terms of traffic congestion. This chapter also examines the potential to allow a greater range of people to drive, and the possibilities opened up by new and emerging technology.

Chapter 3 provides background information and a literature survey. Since the work presented in this thesis is multi-disciplinary, this literature survey draws on work from a range of fields. These include road accident risk sources and factors, risk mitigation (both existing methods and those currently in development such as advanced driver assistance systems, co-operative driving and autonomous vehicles), driver and vehicle modelling, vehicular ad-hoc networks, and the particular simulation tools used in the course of this research.

The risk limit model is detailed in Chapter 4, including determination of current risk levels and vehicle behaviours that can be taken as a result of comparing this level to the risk limit. This chapter also includes an examination of and algorithms for risk-aware link choice and social link choice. Chapter 5 then describes the methodology used for experimental investigations into the feasibility and effectiveness of the risk model and Chapter 6 discusses the results of these.

Chapter 7 describes the design of the networking aspects of the system and algorithms for dealing with these. The main work presented in this chapter is the algorithm for coupled risk estimation, in which vehicles exchange information about accident risk levels and adjust their risk values based on information received from their neighbours. A convergence proof is also provided for this algorithm. The effects of network phenomena on the system are also examined in this chapter, such as the effect of the proportion of vehicles participating in the network, how convergence rate varies with network size and density, and the effects of beaconing rates on the accuracy of the coupled risk estimation algorithm. Chapter 8 describes experiments relating to network performance and the algorithms developed in Chapter 7, while Chapter 9 then provides the results of these experiments.

Finally, Chapter 10 explores questions raised by this research and avenues for further investigation and Chapter 11 concludes this thesis.

Chapter 2

Motivation

The road system is a major part of most people's everyday lives and the problems associated with poor traffic management are well-known. In this chapter, we will discuss these problems and provide concrete data which documents their scope and magnitude. Given the severity of the costs resulting from failures in either road safety or utility, the primary goal in traffic management is to find the best balance between the two in an attempt to minimise these costs overall.

It is not always possible to control this balance. When there is a breakdown in utility, which may be simply from too great a volume of traffic, not only from deficiencies in traffic management, we get congestion. Most people are familiar, on a personal level, with the detrimental effects of traffic jams. From the frustration of crawling along when one is anxious to get home, to the stress of missed appointments or regret of time that could have been better spent engaged in work, leisure or with family, the effects on our everyday lives are far-reaching. On a larger scale, traffic congestion has major economic and environmental impacts, especially when it is widespread throughout a city or region.

The effects of a failure in road safety, however, are even more severe. The human cost of each injury or fatality on the road is devastating for relatives and friends and the economic effects are also substantial: medical bills, lost time at work, damage to property, and often also a road blockage while the accident is cleared. In developed countries, road accidents are one of the biggest causes of death and governments spend heavily each year seeking to control their frequency and severity.

Attempts to balance safety with utility through currently available traffic management techniques unfortunately exclude some from driving altogether. As driving is one of the most common methods of transport, many of our cities are designed to accommodate and even require it. Whilst inner-city areas often have efficient and functional public transportation systems (though this is not universally true), suburbs and regional areas typically do not, so that an inability to drive means an inability to travel, even to conduct the normal activities of everyday life such as going to work or school, shopping, or visiting friends and family. Improved techniques for mitigating accident risk may allow people currently disqualified from driving to be able to do so safely. This would bring a great improvement to these people's lives and lower costs to the community as they gain more independence and rely less on services provided by government or assistance from others.

2.1 Road Safety

According to World Health Organisation data [1], an average of over 850 000 people die in road accidents each year globally, and between 20 million and 30 million are injured. Road accidents are among the top ten causes of death for people aged 5-59 years, and the eleventh most common cause of death overall. For young people the problem is especially critical, with road accidents the second highest cause of death of people aged 5-29. Even when accidents are not fatal, the consequences of injuries can include permanent disability resulting in dependence on others for daily living, chronic physical pain, limitations in physical activity, or permanent disfigurement resulting in emotional trauma.

Even road accidents that do not result in severe injuries or death can nonetheless cause considerable suffering. A study conducted in Sweden [2] found a high rate of psychosocial complications following even minor road accidents. Half the respondents to this study still had travel anxiety two years after the incident, and 16% of those employed could not return to their former jobs. Other commonly reported consequences were pain, fear, fatigue and a reduction in leisure time activities.

In addition to the personal costs, traffic accidents also entail major economic costs. The total global cost of road crashes was estimated to be US$517.8 billion in 2000 [3]. A detailed study in the US found that for the same year, the total national economic cost of road motor vehicle crashes was US$230.6 billion, of which medical costs accounted for US$32.6 billion, property damage US$59 billion, lost productivity (both market and

Type	Cost (USD millions)	% total cost
Medical	$32 622	14.15%
Emergency services	$1 453	0.63%
Market productivity	$60 991	26.45%
Household productivity	$20 151	8.74%
Insurance administration	$15 167	6.58%
Workplace cost	$4 472	1.94%
Legal costs	$11 118	4.82%
Travel delay	$25 560	11.09%
Property damage	$59 036	25.60%
Total	**$230 568**	**100.00%**

Table 2.1: Total economic costs of road accidents in the US for the year 2000. Amounts shown are in 2000 dollars. Totals may not add due to rounding. Reproduced from [4].

household) US$81 billion and other related costs US$58 billion [4]. A breakdown of costs of different types can be found in Table 2.1.

2.2 Congestion

The Texas Transportation Institute's annual Urban Mobility Report for 2011 [5] found that the total delay due to traffic congestion in 2010 in the US was 4.82 billion hours. This resulted in a national total of 1.94 billion gallons (7.34 billion litres) of wasted fuel and an economic cost of US$100.9 billion. Per commuter, on average, this translates to a yearly delay of 34.4 hours and 14 gallons (53 litres) of wasted fuel — equivalent to a week's worth of fuel for the average US driver. In areas with over one million inhabitants, the costs are even greater: 44 hours and 20 gallons (76 litres) of fuel per person. Peak "hour" actually lasted 6 hours in the largest areas.

In addition to wasted time and money, congestion is also detrimental to the environment due to increased levels of greenhouse gas emissions. CO_2 emissions are greatest at low speeds [6], such as are found in congested traffic conditions, and vehicles travelling in such conditions also spend more time on the road. [6] also estimates that congestion mitigation could reduce CO_2 emissions by between 7 and 12 percent in the US.

2.3 The Ability to Drive

Most research on the effects of the inability to drive a vehicle have focused on elderly drivers who cease driving, by choice or otherwise, as this is the largest group of adult non-drivers and also because this allows for comparison between before and after driving cessation. However, many of the concerns raised by this research also apply to younger adults who are unable to drive due to disability or injury.

In [7], a focus-group study was conducted of people over the age of 60 who had stopped driving within the previous two years. Participants in this study fell into two groups: those who proactively chose to stop driving, often because of health concerns, and those who did so reluctantly, for example after failing an eye test or having a frightening experience such as falling asleep while driving. However, participants also discussed a third group, not represented in the study since participants were self-selected and were required to have already stopped driving: those who resist driving cessation even when peers, family and/or doctors consider that they should stop. Thus this study excludes those who might be considered to have the strongest feelings against an inability to drive and yet those who did participate still described profound effects that driving cessation had had on their lives.

One of the main results of driving cessation as reported by participants in this study was a feeling of loss of independence. While some participants had alternative transportation options, many were reluctant to take public transport as it was difficult for them due to physical impairments or because they perceived it as dangerous. Moreover, relying on public transport resulted in a loss of spontaneity for the participants as they needed to plan their trips in advance according to the transit timetable, rather than being able to take trips as and when they needed or wanted to. Most participants also relied at least partly on family or friends for alternative transport but were worried about becoming a burden on them and tried to keep their reliance on others to only trips that were considered a necessity, such as seeing the doctor. This concern may then result in a decrease in leisure or social trips, causing increased social isolation, something that is already a significant concern for the elderly and disabled.

These findings are supported by [8], which used an interview format to investigate the effects of driving cessation on older adults' quality of life. This study also found that participants' independence was compromised and that for some participants, transportation had been reduced to only necessary trips such as doctor's appointments and grocery shopping, rather than for leisure activities — again, participants were reluctant to request family and friends drive them for any trips that were seen as unnecessary. In addition, participants reported having limited friendships. [9] also found that driving cessation was associated with reduced social integration in the form of friendship networks and that this relationship was not affected by the ability to use public transportation.

2.4 New and Emerging Technology

Given the significance and scale of these problems, any method of improving how we manage both the utility and safety of our road system is desirable. This area of research is considerable and many such methods have already been proposed and implemented. However, the advent of vehicular ad-hoc networks (VANETs), allowing communications between vehicles as they travel, is poised to bring profound changes to traffic management in coming years. VANETs, coupled with computational power in vehicles, are set to dramatically alter all aspects of road use. In terms of traffic management, these networks allow for up-to-date, real-time information to be used when making traffic management decisions, both at the level of the individual vehicle and driver, and on a larger scale for traffic authorities. Additionally, the information available can be specific to a location, situation, vehicle or driver. A detailed examination of VANETs and their applications relating to traffic management will be undertaken in Chapter 3.

As we have seen throughout this chapter, failures in both the safety and utility of our road system can have extremely costly effects. However, often improvements to either safety or utility must come at the expense of the other and so traffic management systems attempt to find the best trade-off between these two concerns. With the use of VANETs and the fine-grained, dynamic information they provide, there is the potential for a more precise and responsive means of finding this balance, reducing the problems discussed above.

Chapter 3

Background

This chapter will provide an overview of the fields that affect my research. These fields, particularly the area of road safety research, contain large bodies of existing literature for which a complete survey is beyond the scope of this thesis. As such, this chapter will focus on the intersection of this literature with my work.

Since the core aim of this research is to provide a new means to managing traffic accident risk, we will first examine accident risk itself, its definition and causes and methods for mitigating it — both those that are currently employed as well as those that are still the subject of ongoing research. The results of this investigation into accident risk are then used to inform the risk limit model discussed in Chapter 4, in which causes and factors in accident risk are used to determine the current risk level, and risk mitigation methods are used in determining vehicle behaviour. We will also look at existing work on driver and vehicle modelling. An accurate model of the driver and vehicle can be used both in simulations and theoretical work to understand and evaluate changes to the risk limit model. As the risk model is expanded to incorporate more and more complex risk factors, it is important to be able to predict the effects of particular risk factors or vehicle behaviours.

We will then discuss VANETs [10]: their characteristics, current technology and standards, and their use in traffic management. The behaviour of these networks is an important consideration throughout the work presented in this thesis, particularly in how the risk limit model interacts with the network and is affected by adverse networking phenomena, presented in Part III. Finally, we will provide an overview of the simulation tools used in the experimental parts of my work.

3.1 Accident Risk

Risk as it relates to traffic accidents consists of four aspects: exposure, crash probability, injury probability and injury outcome [1]. Exposure refers to the amount of use of the system by a user or class of users, i.e. the amount of time spent or distance travelled on the road. Crash probability is the likelihood of being involved in an accident, given a particular exposure. Injury probability is the likelihood of sustaining an injury when involved in an accident and injury outcome is the eventual result of this injury.

Since we are interested in balancing accident risk with road system utility, we will concern ourselves primarily with crash probability and, to a lesser extent, injury probability as these two aspects of risk — and of mitigating risk — have the most influence on utility. The other two elements of risk are beyond the scope of this work. Measures to affect exposure are typically too long-term to be relevant to research focusing on dynamic traffic management through the use of VANETs — their effects will persist and be stable over long periods of time rather than varying with the traffic situation, and cannot be influenced by individual vehicles or drivers as they travel. Injury outcome is largely a problem of logistics, economics and medical science and thus the benefits of a VANET are also limited or non-existent for this cause. Hence for this work, we consider risk mainly in terms of crash probability, as this is the aspect of risk that is most amenable to influence from a VANET-based traffic management system.

In the following subsections, we will first examine factors in accident risk — what causes risk and to what extent each of these factors affect crash probability. We will also discuss some of the challenges involved in modelling and using these risk factors in a traffic management system. From there, we will outline risk mitigation strategies and techniques, both those already in use and those that are based on new and emerging technologies which are still in development.

3.1.1 Accident Risk Factors

A number of studies have been conducted on traffic accidents and their causes. These studies have consistently found that most traffic accidents — percentages found vary between 65% and 75% — are caused by human factors, with the remainder accounted for by either vehicle factors or environmental factors [11–13]. We are particularly interested in those factors which vary between individual drivers, vehicles, or situations as these have the potential to provide scope for improvements in utility by using risk differences to inform vehicle behaviour. In the following subsections we will discuss each of these sources of risk in turn.

3.1.1.1 Human Factors

A substantial review of human factors in traffic accident risk was undertaken in [12]. The authors introduce the term "differential accident involvement" to refer to the variation in accident risk amongst individuals. They examine a wide range of factors that can potentially contribute to higher differential accident involvement and review a number of studies to determine the contributions of these factors. Reaction time and vision factors do not appear to have significant effects on accident proneness, while factors that do include selective attention, field dependence, "life events" such as divorce or financial difficulties, emotional stress and temporary physiological factors such as fatigue and the influence of alcohol. We will examine some of these factors in more detail below.

Another important consideration in [12] was the perceived control of the driver. Drivers who felt that they had lower control over the situation or their vehicle felt themselves to be at higher risk and hence drove more cautiously to compensate, resulting in a lower incidence of accident involvement. It is important to note that it is the perceived level of control that is important, not the actual level. A driver's perceived control can be characterised by their expectations: if the situation consistently matches their expectations, they feel themselves to be in control, however, if it does not, they feel out of control. Thus, driver performance could be conceivably improved by modifying the environment to give drivers a more accurate sense of their control of the situation, so that they will behave cautiously when appropriate.

In [14], a comprehensive review is given of research relating to how vision affects driving. It covers impairments both to the eyes and to the visual processing areas of the brain. Interestingly, raw visual acuity is not highly correlated with accident involvement. Rather, factors such as contrast sensitivity as measured by either an Embedded Frames Test or a Rod and Frame Test, visual attention and field of view proved to be more important to the safe performance of the driving task.

Although in a critical situation a driver's reaction time determines the time taken for the vehicle to stop or evade a hazard, there has been no correlation found between reaction times as measured in the lab and accident proneness in drivers [12,13,15]. This includes both simple and choice reaction time, where the subject must not simply react to a stimulus, but choose a correct reaction, e.g. pushing one of two buttons. In addition, [16] investigates whether there is a relationship between variability of reaction time in an individual and accident rate and found no statistically significant relationship. It has been suggested that this lack of results relating reaction times and accident involvement is due to drivers compensating for slower reaction times and adjusting distances to other vehicles accordingly [12].

Selective attention — the ability to identify and focus on one stimulus in the presence of multiple conflicting stimuli — has been shown to be correlated with accident involvement in a study of 117 bus drivers [17]. The drivers were given a selective attention test involving listening to two conflicting auditory inputs, one in each ear, and repeating only one of them. Errors on this test, in particular on the second part of the test which involved switching attention between the two inputs, correlated with poor accident ratings as given by the bus company based on accident reports, which were mandatory for all drivers.

[18] breaks learning of the driving task into three stages: cognitive, in which thought must be given to all actions, associative, where some but not all actions are automated, and autonomous, where driving itself is largely automatic and cognitive effort is focused on higher-level goals such as navigating to a destination. This paper then identifies causes of accidents in terms of cognitive states such as competing motivations, inexperience or lack of knowledge and established improper and wrong habits.

[19] analyses driver behaviour and mental and physical state prior to traffic incidents — where an incident is defined as an accident or near-accident — in order to identify states and behaviours that contribute to accidents. Some of the behaviours which were identified include desultory driving — not paying proper attention to the road due to talking, drowsiness, etc; no safety confirmation — failing to check that it is safe to perform a maneuvre before doing so, for instance, failing to check for the presence of vehicles in a target lane during a lane change; and inappropriate assumption — failure to accurately make predictions about the driving situation. Physical states identified included haste, lowered concentration and drowsiness. The effects of these states and behaviours were analysed using interviews with 35 participants about their traffic accident histories. The accidents were broken down into seven types and the states and behaviours were ranked in terms of their contributions to the different types of accidents.

Driver age plays a significant role in differential accident involvement. Both young (under 25) and old (over 75) drivers have been shown to have higher accident rates [20, 21]. However, the effects of age on differential accident involvement are complex and do not really correspond to one single cause, making age problematic as a risk factor to be included in a model for traffic management even though it is very easy to test for. In young drivers, driving inexperience, the effects of alcohol, and having passengers in the car have been shown to particularly increase risk [20]. Thus it is not clear whether a driver model should incorporate these factors separately or take age as one, over-arching risk factor.

Similar concerns apply to driver gender. Males, especially younger men, have higher crash rates than women, even when corrected for differences in exposure [21]. However, again, gender represents a constellation of other factors rather than one single risk factor, making it difficult to model as a singular entity.

The two more temporary, situational human factors that play the largest role in accident risk are alcohol consumption and fatigue. There has been research going back to the 1960's [22] (cited in [1]) showing that drivers who have consumed alcohol have a higher risk of accident involvement than those with zero blood alcohol content (BAC), and that this risk increases with BAC. Alcohol is a significant risk factor even at relatively low levels. A large case-control study [23] found that the relative risk of crash involvement starts to increase significantly at a BAC of 0.04 g/dl.

A population-based case-control study of 571 drivers involved in crashes, along with 588 control drivers [24], found that the population-attributable risk for driving with at least one of three measures of sleepiness was 19%. Crashes used in this study were only those where at least one occupant was admitted to hospital or killed. The fatigue indicators used were driving while feeling sleepy, driving after less than five hours sleep in the preceding 24 hours and driving between 2 am and 5 am.

3.1.1.2 Vehicle Factors

Apart from actual mechanical failures or defects, the effects of vehicles themselves on road safety largely involve various risk mitigation features and technologies. While the absence of these may be considered a risk factor, we have deferred a discussion of them to Section 3.2, where we consider techniques for risk mitigation.

Accident risk is also affected by the physical properties of the vehicle, such as its mass, size and maximum deceleration. These primarily affect accident risk through their interaction with vehicle speed, in particular because they affect the stopping distance of the vehicle at any given speed. Vehicle speed may perhaps be considered a combined driver and vehicle factor, since although it is the speed of the vehicle in question, it is the driver who chooses and controls this speed. However, most driver factors relate to the driver's characteristics and state and speed is more properly considered part of the state of the vehicle, so we discuss it here.

Vehicle speed is one of the most significant and well-studied factors in accident risk and is frequently cited as a contributing factor in accident reports. The probability of a crash involving an injury has been found to be proportional to the square of the speed, while the probability of a fatal crash is proportional to the fourth power of speed [25]. Speed studies in various countries show an increase of 1 km/hr in mean traffic speed typically results in a 3% increase in injury crashes and a 4–5% increase in fatality crashes [26]. A meta-analysis [27] of 51 studies on speed limit changes fitted a model to the relationship between speed changes and accident rates, in which the change in accidents with respect to speed is exponential, given by

$$\frac{\text{Accidents after}}{\text{Accidents before}} = \left(\frac{\text{Average speed after}}{\text{Average speed before}}\right)^{\text{Exponent}}$$

with exponents of 3.6 for fatal accidents, 2.4 for accidents with serious injuries and 1.2 for accidents with slight injuries. Speed also has a significant impact on the number and severity of injuries when crashes do occur [1].

Additionally, speed variance between vehicles has previously been put forward as a factor in accident risk, however later research suggests that these results may instead be due to the relationship between speed and crash incidence, rather than speed variance itself [28].

3.1.1.3 Environmental Factors

Environmental factors are those that are external to the driver and vehicle in question but which affect its risk level. Included in this category are factors such as weather, time of day, traffic conditions, and factors relating to the road itself or the area it passes through.

One effect of environmental factors is their impact on visibility of other vehicles. Visibility plays a key role in three types of crashes: a moving vehicle running into a slowly moving or stationary vehicle located ahead at night time, angled and head-on collisions in daytime, and rear-end collisions in fog (at all times of day) [29]. Similarly to driver age and gender, however, visibility may itself be made up of a combination of other factors, such as weather (especially fog), time of day, and road geometry and features.

Adverse weather such as rain, ice or snow may affect the friction coefficient of the road surface in addition to reducing visibility, further increasing accident risk [30, 31]. [31] calculates accident risk factors for 10 different types of road slipperiness, with these values ranging from 1.5, for drifting snow or hoarfrost coupled with low visibility, up to 11.6 for precipitation — rain or sleet — on a frozen road surface. The non-slippery condition had an accident risk factor of 0.7. Moreover, drivers tend to judge poorly the adjustments to driving behaviour, such as reduced speed, that are required to compensate for slippery road surfaces [30].

In addition to their role in affecting visibility of other vehicles, road type geometry and other features also affect the manoeuvring of the vehicle and make a significant contribution to accident risk [32, 33]. In particular, roadside hazards contribute to between 25% and 40% of fatal crashes [34]. Some methods for reducing this problem will be discussed in Section 3.2.

Road planning can also affect accident risk by modifying the environment through which the road passes. Some risk factors relating to road planning are through-traffic passing through residential areas, conflicts between pedestrians and vehicles near schools located on busy roads, lack of segregation of pedestrians and high-speed traffic, lack of median barriers to prevent dangerous overtaking on single-carriage roads and lack of barriers to prevent pedestrian access onto high-speed dual-carriageway roads [35].

3.1.2 Measuring Risk

Separate from the identification of factors contributing to accident risk is the problem of testing for and measuring these factors. Risk factors relating to the vehicle and environment are typically grounded in comparatively easily measurable characteristics such as physical mechanics, the presence of various types of hazards and current conditions such as weather and time of day. In contrast, risk factors relating to the driver can often be much harder to measure and thus require further examination. Some of the factors above, such as selective attention and field dependence, have well-defined tests to determine where a particular individual lies on the continuum. Other factors, however, especially those relating to the internal mental state of the driver, are harder to test for. It is also unclear how a driver's scores for many separate risk factors can be best combined to produce a single dimension of risk, given that each factor alone does not determine whether a driver is safe or unsafe and that many factors can interact to produce unexpected results.

A range of tests intended to predict safe driving in individuals were reviewed in [36]. Some of these are tests or batteries of tests for one or more specific risk factors, while others are more generally targeted towards performance on the driving task. The tests were examined in terms of their specificity and sensitivity. Specificity refers to the test's ability to correctly categorise drivers who are safe, i.e. a high rate of true negatives, whereas sensitivity refers to the test's ability to correctly identify unsafe drivers, i.e. a high rate of true positives. These two measures together define the usefulness of each test, that is, whether it is capable of accurately separating drivers into two groups: safe and unsafe. It was found that most existing tests did not have high rates of both specificity and sensitivity and hence were not accurate in predicting which drivers were at risk. Those that did were tests that only applied to drivers with certain medical conditions and hence not as useful in providing risk estimations for all drivers.

In response to the need for a test which is applicable to a wide range of drivers and has high levels of sensitivity and specificity, the DriveSafe test (formerly the Visual Recognition Slide Test) was developed and evaluated in [37]. The evaluation measure used was an on-road assessment of the drivers tested, which was validated in [38]. In the DriveSafe test, participants are shown scenes depicting a rotary intersection with a number of pedestrians or other vehicles present, and then asked to recall the positions and orientations of these. This test was shown to have a sensitivity of 81% and a specificity of 90%. However, this was further improved by the addition of DriveAware, a questionnaire which measures drivers' awareness of their driving ability [36,39]. By combining both these tests, a specificity of 96% and sensitivity of 95% was achieved [40]. These tests would thus provide a functional basis for determining the risk level of individual drivers.

3.2 Mitigating Risk

We will now examine methods for mitigating accident risk. We will begin by discussing methods that are already well-established and widely used. From there, we will move on to risk mitigation methods based on emerging technologies.

3.2.1 Current Methods

Methods for mitigating accident risk that are currently employed tend to be clustered around three main areas: managing exposure to risk, safety-awareness in planning and designing roads, and safer vehicles [1]. Strategies for managing risk exposure tend to be very long-term, for example, land use policies, design of overall road networks, or licencing policies. Hence, they are of less relevance to this thesis than the other two methods, which are more immediate and more easily changed over short periods of time. Since we are concerned with determining the risk level for an individual vehicle and driver in a particular situation, the influence of large-scale, long-term factors such as land use policies is not as important or relevant. Instead, we will focus on the latter two areas of safer roads and safer vehicles.

One of the most widely-used means of controlling accident risk on roads is the use of speed limits, and in particular, setting speed limits that are appropriate to road functions. For example, a motorway will typically have a higher speed limit than a small residential street. In the Netherlands, implementing a system of speed limits assigned based on road functions resulted in a reduction of more than a third in the number of injury crashes [41].

However, this finding contrasts with an examination of speed limit laws in the United States. In 1974, the US federal government enacted a law to limit speed on federal highways to 55 mph (88.5 km/hr) in order to save fuel. This law was subsequently altered to an optional 65 mph (104.6 km/hr) in 1987 and then repealed entirely in 1995. In [42], the authors study the effects of the 1995 law changes, which varied from state to state, with some states keeping the 65 mph limit, others raising it and one state, Montana, removing daytime limits on federal highways entirely. After taking into account various factors such as seasonal changes in driving patterns, unemployment rates, etc., most changes in fatality rates were not statistically significant. One possible factor reconciling these differing results is that rates of non-compliance with the speed limit on federal highways were very high, especially after the introduction of the 55 mph limit [42].

Road layout and design can also be utilised to minimise accident risk. Motorways and other high-speed roads typically have a combination of some or all of: large-radius curves (both vertical and horizontal), "forgiving" roadsides that do not have hazards or are designed to minimise damage in the event of a collision, grade separated junctions, and median barriers. Because of these features and the fact that non-motorised traffic is disallowed, such roads have the lowest rate of road injury per distance travelled [43].

It is often not feasible for smaller, rural roads to have these features but they can nonetheless be improved in terms of road safety. This can be achieved by measures such as making provisions for slow and vulnerable road users, overtaking and turning lanes, median barriers, improved vertical alignment, regular speed limit signs, advisory speed limits at sharp bends, and rumble strips [1]. In addition, accident hazards can be reduced by road lighting to highlight hazards or intersections and the removal of roadside hazards such as trees or telegraph poles [1].

Areas which are shared between motorised vehicles and other traffic such as bicycles or pedestrians present their own particular forms of accident risk, for which risk mitigation methods in shared areas primarily involve traffic slowing or calming measures [1]. Examples of these include roundabouts, chicanes, road narrowings, speed bumps, and preventing motorised vehicles from entering certain areas, leaving these for bicycles and/or pedestrians alone. These measures have achieved crash reductions of between 15% and 80% in Europe [1]. More general measures that apply to a variety of different road types include preventing road use that does not match the intended function of the road, separating different kinds of road users (for example by having dedicated bicycle lanes), and making it clear through signage or otherwise what is and is not appropriate road use.

Roadsides are a prime candidate for risk mitigation measures as many accidents involve collisions between vehicles and roadside objects: in the US, approximately one-third of all highway fatalities involve a collision with a roadside object [44]. Some measures to reduce the impact of these accidents include avoiding cut side slopes, decreasing the distance from outside shoulder edge to guardrail, decreasing the number of isolated trees along roadway sections and increasing the distance from outside shoulder edge to light poles [44]. Additionally, US Transportation Research Board evaluations showed crash cushions reduced fatal and serious injuries by up to 75% [45].

In terms of improving the safety of vehicles themselves, there are two major categories of risk mitigation measures. These are improvements to the visibility of vehicles, which are aimed at preventing accidents, and crash-protective vehicle design, which are aimed at reducing harm to occupants of the vehicle in the event of a crash [1]. Two significant improvements to vehicle visibility are daytime running lights — which lead to a reduction in crashes of between 8% and 15% — and high-mounted stop lights, which lead to a reduction of between 15% and 50% in rear crashes [1].

Crash-protective vehicle design consists mainly of safer car fronts to protect pedestrians and cyclists, seatbelts, airbags, and frontal and side impact protection, which prevents any intrusions into the interior of the vehicle during a crash and thus allows seatbelts and airbags to operate correctly. It has been estimated that take-up of testing standards for vehicle fronts developed by the European Enhanced Vehicle Safety Committee could avoid 20% of deaths and serious injuries to pedestrians and cyclists in EU countries annually [1]. Seatbelts have been found to reduce the risk of serious and fatal injuries by between 40% and 65% [1], while airbags, when combined with seatbelt use, reduce the risk of death in frontal crashes by 68% [46].

In addition to designing vehicles that lessen the impact of an accident, there are a number of vehicle features which can be employed to improve vehicle conspicuity in order to prevent collisions from occurring in the first place. Daytime running lights, reflective areas of vehicles (such as licence plates) and high-mounted brake lights have all been shown to reduce accidents, particularly in low-visibility conditions. [29].

All of these existing risk-mitigation measures, while effective, are static and therefore do not adapt to different or rapidly-changing situations. They must always be calibrated to the worst possible case, since it is not currently possible to know the circumstances at any given time, or the risk factors for a given driver or vehicle. As a result, they are not optimal for balancing risk with utility; in situations where the overall risk is lower, all of the above risk mitigation methods will still be in place, including those that have a significant negative impact on utility, such as speed limits.

3.2.2 Emerging Technologies

We will now examine risk mitigation methods which use technologies that are still in development. In particular, we will consider advanced driver assistance systems (ADASs), co-operative driving and autonomous vehicles.

3.2.2.1 Advanced Driver Assistance Systems

Advanced Driver Assistance Systems (ADASs) are electronic systems to help the driver of a vehicle with the task of driving. This help might take different forms, such as navigation assistance, improved driver comfort, improvements to safety either in terms of preventing accidents or lessening their effects, or improvements to road system utility. There have been multiple different methods of classifying and analysing ADASs, often based on the function they perform, the technology used for implementation, or vehicle or road type [47], however here we will focus on a safety and utility perspective on ADASs.

Evaluations of the effects of ADASs on road safety and utility are not straightforward since available quantified analysis of these systems is limited — large-scale implementations have not yet taken place, and large-scale on-road experiments are often not feasible to conduct [47, 48]. [47] develops criteria for analysing the safety and road system efficiency of various ADASs, however to overcome the lack of concrete data on ADAS safety outcomes, expert judgement is used to assess the systems. Two main categories of ADAS are considered: driver support systems — including functions relating to driver information, driver perception, driver convenience, and driver monitoring — and vehicle support systems, including systems for general vehicle control, longitudinal and lateral control, and collision avoidance.

The criteria used for road safety are avoidance of inappropriate speed, keeping appropriate longitudinal and lateral distances, and support of driver awareness. While these criteria represent common accident causes and thus systems which have an effect on one or more of the criteria can be expected to reduce accident rate, this is not a direct measure of the safety impacts of ADASs. Moreover, there is a question of what are considered “appropriate” values for speed and distances, especially in the context of a system which takes a more dynamic and situational approach to risk mitigation, such as developed in this thesis, as these values would generally be variable in such a system. However, using an ADAS to control these values, once set, would allow for less conservative values to be used, as an ADAS can be typically expected to maintain desired speeds or distances with a smaller margin of error than is required by a human driver.

The criteria used for traffic efficiency are speed adjustment and headway adjustment. These criteria work well as one commonly-used measure of road system utility is average vehicle speed, and another is traffic flow, which can be derived directly from the combination of speed and headway. However, broader congestion control measures such as vehicle routing are not included.

Despite these limitations, the assessment of a large number of ADASs according to the criteria outlined above (shown in Table 3.2) does provide a good overview of the expected safety and efficiency impacts of implementing such systems. One caveat is that in this work the negative effects of ADASs are assumed to be minimal, which may not in fact be the case, as the driver distraction effects of ADAS interfaces are still an area of ongoing research [49].

In [50], the expected safety impact of various ADASs is compared with their response level, with low-response systems being those that have a response time greater than a driver's reaction time, and high-response systems those with a response time faster than a driver's reaction time. A collision avoidance system which automatically applied the brakes, for example, would fall into this latter category. ADASs were plotted on a comfort-safety axis against their response time. In general, ADASs which were considered to have a greater potential impact on safety were also high-response systems, with decreasing response time correlated with increasing safety benefits, with the only exception being a wrong-way driver information system, which was low-response but had high impact on safety. This trend suggests that in order to see the full benefits of ADASs in terms of road safety, it will be necessary to allow them more and more direct control of the vehicle, rather than acting through the driver as an intermediary.

Another methodology for estimating the potential safety impacts of ADASs, used in [48], is to compare the function of each ADAS to a specific accident cause. Causes are usually recorded in traffic authority databases and so there is a large amount of data available for analysis. Matching of ADASs to accident causes is possible because each ADAS will typically be specific to a particular driving task, relating directly to one accident cause. However, this relies on the assumption that the ADAS would be operational 100% of the time and ignores multiple-cause accidents. Nonetheless, this provides a good basis for comparing the relative effects of different ADASs. In addition to ADASs relating to specific accident causes, intelligent speed adaptation was also analysed. However here, a different methodology was required, since speed is cited as a contributing factor in most accidents. Instead, a correlation between speed and accident rate [51] was used to estimate the reduction in accident rate that could be expected from widespread adoption of this system. The results of this analysis are shown in Table 3.1.

In addition to these broader studies, some research has been done on the expected

Speed limit	100–120 km/hr			50–90 km/hr			30–50 km/hr		
ADAS	Fatal	Heavy injury	Slight injury	Fatal	Heavy injury	Slight injury	Fatal	Heavy injury	Slight injury
Speed headway keeping	14%	21%	35%	4%	7%	18%	—	—	—
Front obstacle collision avoidance	28%	13%	14%	26%	20%	18%	29%	26%	22%
Lane keeping support	35%	36%	24%	33%	25%	16%	12%	9%	7%
Side obstacle collision avoidance	2%	9%	11%	1%	1%	2%	2%	2%	3%
Intelligent speed adaptation	10%	8%	—	30%	24%	—	38%	30%	—

Table 3.1: Projected relative accident rate decreases (%) with widespread use of various advanced driver assistance systems in the Netherlands, from [48]. Empty cells indicate cases where data was not available.

safety benefits of specific ADASs. [52] provides recommendations for implementing intelligent speed adaptation in the UK, including estimates of the proposed implementation's effect on accident rates. These estimates are based on a model of accident rate as a function of mean speed developed in [25]. Figures calculated for reduction in injury accidents range from 10% for an advisory system up to 20% for a mandatory system which the driver cannot override. (Figures cited are for fixed speed limits.) For fatal accidents, the reductions calculated range from 18% for the advisory system to 37% for the mandatory system.

[53] uses induced exposure methods [54, 55] (see Section 4.2.1 for further discussion of this methodology) to match accidents in Sweden during 2000 to 2002 that would be sensitive to the use of an electronic stability program that would maintain the stability of a vehicle in low-friction conditions such as wet or icy roads. The overall effectiveness of this system, i.e. the proportion of accidents which the system would potentially prevent, was estimated at 22.1%. Restricting the analysis to accidents on wet roads gave effectiveness of 31.5%, while for icy roads, the effectiveness was calculated as 38.2%.

While concrete, real-world data on the safety effects on ADASs are not yet available, studies such as these point the way towards significant impacts on accident rates were these systems to achieve widespread implementation. As such their presence in a vehicle is important to include in a model for risk mitigation and traffic management in the future. For this purpose, most ADASs effectively act as inverse risk factors, reducing the likelihood of an accident rather than increasing it. The presence of these systems may allow a vehicle or driver more leeway to employ utility-increasing behaviours since the ADAS(s) would help in controlling the risk level, thus making risk-mitigating behaviours less necessary and allowing a greater range of utility-increasing behaviours.

One factor limiting the deployment of ADASs is that high-response systems — those that respond rapidly to a situation and take actions without the driver's input — require highly reliable technology for sensing and understanding the traffic situation [48]. While this is not yet widely available, low-response ADASs are nonetheless a step towards more autonomous control of vehicles, a path that will be furthered by the advent of high-response systems, leading eventually towards fully autonomous vehicle control. As such, future traffic management systems must be able to accommodate varying levels of autonomous vehicle control and a heterogeneous mix of vehicle control types on the road. The model developed in this thesis is compatible with this heterogeneity as the

			Road safety			Traffic efficiency	
			Avoidance of inappropriate speed	Keeping appropriate longitudinal and lateral distance	Supporting driver awareness	Speed adjustment	Headway adjustment
Driver	Driver information	Navigation routing	L	L	L	L	L
		Integrated navigation	H	L	L	L	L
		Real-time traffic and traveller information	L	L	L	H	L
	Driver perception	Vision enhancement	L	H	L	H	L
		Electronic mirror	L	H	L	L	L
		Parking and reversing aid	L	L	L	L	L
		State of the road surface system	H	H	H	H	L
	Driver convenience	Driver identification	L	L	H	L	L
		Hands-free and remote control	L	L	H	L	L
		Automated transactions	L	L	L	H	H
	Driver monitoring	Driver vigilance monitoring	L	L	H	L	L
		Driver health monitoring	L	L	H	L	L

Vehicle	General vehicle control	Automatic stop and go	H	H	L	L	L
		Platooning	L	L	L	H	H
		Speed control	H	H	L	L	L
		Adaptive cruise control	H	H	L	H	H
	Collision avoidance	Road and lane departure collision avoidance	L	H	L	L	L
		Lane change and merge collision avoidance	L	H	L	L	H
		Rear-end collision avoidance	L	H	L	L	L
		Obstacle and pedestrian detection	H	L	H	L	L
		Intersection collision warning	H	L	H	L	L
	Vehicle monitoring	Tachograph	L	L	L	L	L
		Alerting systems	L	L	L	L	L
		Vehicle diagnostics	L	L	L	L	L

Table 3.2: Road safety and traffic efficiency impact of various advanced driver assistance systems, from [47], L = low impact, H = high impact

presence of various types of ADAS can either be incorporated as risk factors (in this case ones that decrease risk) and/or provide additional vehicle behaviours for either mitigating risk or increasing utility.

3.2.2.2 Co-operative Driving

Related to advanced driver assistance systems is the concept of co-operative driving: the co-ordination of multiple vehicles through the use of inter-vehicle communications (IVC). Applications of co-operative driving have thus far been mainly focused on improving the efficiency of the road system with techniques such as co-operative adaptive cruise control and platooning, intelligent merging, emergency vehicle prioritisation, and intersection management. These will be discussed in more detail in Section 3.4.3.

Here we focus on the safety aspects of co-operative driving, which have so far been fairly limited, with the main application being collision warning systems [56–59]. These can be divided into two main types: highway collision warning systems and intersection collision warning systems. Highway collision warning systems allow vehicles travelling in the same direction to be warned of impending rear-end collisions. In intersection collision warning systems, vehicles are warned of potential collisions with other vehicles travelling with an intersecting trajectory, which may not yet be visible to the driver [60].

While collision warning systems have the potential to prevent many accidents, this technology is still in its infancy and most of the research effort has so far been focused on technology and implementation rather than analysis of the safety impacts of such systems [60, 61]. The extent to which co-operative driving will affect accident risk and road safety is thus very much an open question. However, an attempt was made in [62] to quantify the upper limit of the impact of vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) systems on accident rates. This study found that V2V systems could potentially address up to 79% of accidents, V2I systems up to 26% of accidents and both combined up to 81%. However, as this is the maximum possible theoretical impact of such a technology, real world performance would likely be lower, even with 100% penetration of the system.

3.2.2.3 Autonomous Vehicles

An autonomous vehicle is one which is capable of driving without a human controlling it. Research in this field has been active since the 1980's but has only recently matured to the point of fully operational on-road implementations. Selected representative research in the extensive domain of autonomous vehicle development is presented here, highlighting elements relevant to the scope of this work.

The DARPA Grand Challenge events [63, 64] were competitions intended to stimulate research in the field of autonomous vehicles, run by the Defense Advanced Research Projects Agency (DARPA) in the USA. In these events, teams entered vehicles which were required to drive autonomously across desert terrain along a predefined route. This was then further developed into the DARPA Urban Challenge [65], in which vehicles were required to drive autonomously in an urban setting involving both manned and autonomous traffic. For all of these challenges, teams from around the world competed. In the Urban Challenge, 89 teams applied initially, with 11 of these making it all the way through to the final event.

The VisLab Intercontinental Autonomous Challenge [66] ran in 2010 and was a test of autonomous driving along a route of more than 13 000 km from Italy to China. This trip took over three months to complete and no global path planner was used due to the unavailability of maps covering the entire route. Instead, vehicles travelled in lane-keeping, waypoint-following or follow-the-leader mode to facilitate navigation. This test involved varied and challenging terrain at an average speed of 38.4 km/hr (the maximum speed reached was 70.9 km/hr) and provides a large-scale demonstration of the capabilities of autonomous vehicles.

Perhaps one of the best-known autonomous vehicles is being developed by Google [67]. Their driverless car has logged over 140 000 miles (225 308 km) of autonomous driving with no at-fault incidents to date. Notably, lobbying from Google has resulted in the Nevada Legislature passing a law in June 2011 allowing autonomous vehicles to obtain licences to drive on Nevada roads [68]. The Google driverless car, a modified Toyota Prius, became the first vehicle to receive such a licence for testing on public roads in May 2012.

A number of vehicle manufacturers are also investigating autonomous vehicles and related technologies, including Audi [69], General Motors [70], Volvo [71], and Volkswagen [72]. Many of these projects do not aim yet at fully autonomous and ubiquitous control, instead limiting themselves to control at low speeds in traffic jams, or systems similar to co-operative adaptive cruise control. However, these projects represent an increasing trend towards implementing autonomous control in commercially-available vehicles.

Since self-driving vehicles are still a developing area of research, with very limited deployment to-date, there has not yet been much research into the safety impacts of a road system populated by autonomous vehicles. However, since human factors account for a large percentage of accidents (see Section 3.1.1), self-driving vehicles could potentially have a huge impact on road safety simply by removing the driver from the equation. Autonomous vehicles eliminate human error from the vehicle control task (excluding, of course, errors made by the humans programming the vehicle) and remove risk factors relating to human biology such as fatigue, age, gender, blood alcohol level, and cognitive or physical impairments.

Beyond this, in fact, autonomous vehicles may have enhanced capabilities over a human driver. Many sensors used in self-driving vehicles have better range or acuity, or qualitatively different perception capabilities than human senses and, unlike for a human, the task of processing information from all the sensors can be distributed to multiple processors, thus not distracting from the core task of driving as more sensors are added. This also applies to communications between vehicles: while for a human driver, designing the user interface for presenting information received over the network is difficult, this is not an issue for an autonomous vehicle. While there are still challenges to be addressed in areas such as visual information processing and real-time algorithms for vehicle control and route planning, both the hardware and software of an autonomous vehicle can be upgraded as technology improves.

The advent of autonomous vehicles has the potential to make the development and implementation of traffic management based on risk limits significantly more feasible as it addresses some of the major challenges involved. In particular, widespread adoption of autonomous vehicles would abrogate the need for driver modelling, which is a particularly difficult aspect of implementing any system based on risk limits. Additionally, issues such as driver acceptance and compliance and user interface design are avoided (for further discussion of these, see Chapter 10).

The concept of risk limits remains relevant for a road system involving autonomous vehicles. Even though human factors account for a large proportion of accident risk, there is still significant differentiation in risk between vehicles, environments and circumstances that can be exploited to improve utility. These differences arise from factors such as weather, different vehicle capabilities and safety features, traffic conditions, and so on. In particular, the transition to ubiquitous use of autonomous vehicles is not likely to be sharply delineated but rather a gradual decrease in direct human control with increasing use of advanced driver assistance systems, and involving a period of mixed autonomous and human-controlled make-up of vehicles on the road. Hence there will be a need to manage this heterogeneous road environment and using risk limits for traffic management may be one way to accomplish this, as this system is based on risk differentiation and thus equipped to handle a variety of differing capabilities gracefully.

3.3 Driver and Vehicle Modelling

Existing research into driver and vehicle modeling can be characterised by its aims. There is currently a strong level of interest in advanced driver assistance systems (ADASs have been discussed in more detail in Section 3.2.2.1). A number of these systems incorporate a driver model to some degree in order to inform the operation of the system. These models tend to be concerned with the immediate state of the driver, such as where their attention is currently focused or whether they are currently fatigued.

Another category of driver models which take a more long-term view are those of a more cognitive nature. These models are concerned with determining how a human performs the driving task: the perception, thinking and internal processes that transform external stimuli, such as the view of the road, into actions, such as turning the steering wheel or applying the brake.

Lastly, there are models which focus solely on the behaviour of the driver and vehicle. These models are not as focused on understanding the internal processing that occurs as cognitive models are but are rather interested in its external results: the actions drivers perform in a given situation. Because of this behavioural focus, these models can be readily used in simulations and to understand how both microscopic and macroscopic vehicle and traffic phenomena are produced.

3.3.1 ADAS Models

One example of the first type of model can be found in [73]. In this study, computer vision techniques are used to monitor the driver's eyes in real-time in order to detect events such as the eyes closing. This allows the system to determine if the driver is becoming drowsy or has fallen asleep and give a warning. The system was shown to work in real-time with greater than 90% accuracy.

[74] and [75] are both aimed at gathering information about the environment to determine the current driving situation and the current maneuvre — such as a lane change — that the driver is performing. [75] is concerned specifically with identifying driver behaviour such as indicator use, steering wheel position and eye movement, which precedes a lane change maneuvre, in order to automatically detect the driver's intention to change lanes. [74] uses a Bayesian network to identify different driving situations from information about the environment — such as the positions of other cars on the road, and the current vehicle state — such as steering and braking. In particular, [74] focuses on detecting emergency braking situations for use in an ADAS to provide automatic braking in these situations when the driver fails to.

[76] describes an active speed management system that adapts to individual driving styles. Drivers first use a driving simulator so that the system can determine their driving style, which includes parameters such as preferred deceleration rates for corners and intersections and preferred speeds for given road types. The system will then determine appropriate target speeds for each driver and give feedback by use of both visual displays and a haptic accelerator pedal, so that when drivers are approaching or exceeding the target speed, acceleration becomes more difficult. A number of environmental factors are taken into account, such as the current speed limit, road type, presence of other vehicles and pedestrians, and traffic control devices such as stop signs or traffic lights.

Although these systems do not give an overall model of a driver, they are successful in modeling one or more specific parameters and using them to assist with the driving task. These systems also give an indication of the processing required to determine appropriate behaviour from the real-time stimuli associated with driving, and examples of how that processing can be implemented. The main drawback of ADAS driver models, in terms of developing a general model for risk mitigation and traffic management, is that they are very specific, focusing on only one aspect of driver behaviour in a single vehicle, rather than a more holistic view of the driver and vehicle, or even of groups of vehicles.

3.3.2 Cognitive Models

In [77] and [78], the ACT-R cognitive modeling architecture is used to create a driver model for highway driving, including steering, lane-keeping and lane changes. In this model higher-level decision making drives lower-level control and monitoring behaviours, such as where to focus attention, steering angle, and acceleration and braking. The model was tested by comparing the behaviour it produced to observed human behaviour in a driving simulator, which the model was able to interact directly with, resulting in data that could be directly compared to the data obtained from the human participants. This approach provides a good framework for driver modeling as it allows for refinement of the model by addition of new parameters or by changing individual modules.

In [79], the emotions of drivers are modelled to produce a more realistic view of their behaviour. [80] builds on this by also including personality in the model. Emotional responses are calculated from events in the traffic system and incorporated with an initial emotional state that represents the driver's emotions prior to beginning driving. The Ortony, Clore and Collins model of emotions [81] is used in both [79] and [80] and in [80] the Big 5 model of personality is used as well. Events and objects are rated on a number of factors such as familiarity and desirability and drivers are attributed goals of three types: a-goals are goals being actively pursued such as reaching a destination, i-goals are interest goals such as having good traffic and r-goals are replenishment goals such as the need for more fuel.

[80] uses both the emotional model from [79] and a three-part personality model to map emotions and mood as points on a pleasure-arousal-dominance space in which the emotion is pulled towards the mood point with a force dependent on personality. The personality model is divided based on time; it consists of personality which remains stable over years, mood which remains stable over days and emotions which are immediate and transitory, lasting for minutes. These two studies illustrate how psychological concepts such as emotions and personality can be used in a model of the driving task in a concrete way.

Because these cognitive models capture the internal processing that occurs during the driving task, they have the potential to give a more complete view of driver behaviour than either the highly specific models used in advanced driver assistance systems or external, behaviour-focused models. In particular, cognitive models have a greater predictive value; they have the capability to model not only the current state but also drivers' intentions, and to identify patterns in the behaviour of a driver. They also give insight into how a driver's behaviour might be best modified. However, these models can be harder to implement and require more processing than simpler behavioural models. As such, they may not be as suitable for the implementation of a real-time system.

3.3.3 Behavioural Models

One of the foundational works in the area of driver modelling is [82]. It establishes a mathematical model of a driver and vehicle and treats the driver as a feedback system, in which the desired path is the input and various response variables such as vehicle speed and direction serve as both outputs and feedback to the driver. This paper is mainly concerned with the dynamics of driving, rather than a full model of the driver themselves.

In [83], a model is put forward to capture the car-following and lane-changing behaviour of vehicles. A fundamental assumption of this model is that only the immediately surrounding vehicles are relevant to decisions made about speed, acceleration, following distance and lane-changing. In particular, the acceleration is set based on the current speed and acceleration of the lead car as well as the distance to the lead car. Car following is perceived as a phase space of headway distance and acceleration, and this is broken down into five regions, each of which have particular behaviours associated with them. Lane-changing involves the lead car in the current lane as well as both the lead and following cars in the prospective lane to change into. The lane-changing model presented in this paper is intentionally asymmetric, incorporating the concept of overtaking lanes. This model is also used as the basis for the simulation tool Paramics, which will be discussed in Section 3.5.

For a comparison of the advantages and disadvantages of ADAS, cognitive and behavioural driver models, please see Table 3.3.

ADAS models		Cognitive models		Behavioural models	
Advantages	Disadvantages	Advantages	Disadvantages	Advantages	Disadvantages
Can provide an accurate model of one or more specific aspects of a driver Can make decisions based on driver behaviour in real-time	Do not provide an overall model of a driver	Capture the driver's internal processing and state Good predictive value: can model the driver's intentions and identify patterns in behaviour Can give insight into how best to modify driver behaviour	Hard to implement Require a large amount of processing, may not be suitable for real-time systems	Easy to implement and test	Only model external behaviour, give no understanding of the driver's internal state

Table 3.3: Advantages and disadvantages of different types of driver models

3.4 Vehicular Ad-hoc Networks

A vehicular ad-hoc network (VANET) is a wireless network in which the nodes are vehicles. It is ad-hoc in the sense that network configuration is decentralised and done on-the-fly. Each node can participate in routing and forwarding of data.

VANETs have a number of key differences to other types of mobile ad-hoc networks (MANETs), which stem primarily from using vehicles as network nodes. Often in a MANET, the level of mobility of nodes is much lower than in a vehicular network. Vehicles are constantly moving at high speeds relative to each other and thus passing in and out of transmission range at a high rate, especially when we consider shadowing and fading effects from surrounding buildings in urban environments. This high level of mobility makes traditional routing protocols impractical as the lifespan of a given route is similar to the time taken to discover the route [84]. Additionally, the network topology will necessarily follow the layout of the road network, as most of the time vehicles will be travelling along a road when participating in a VANET. This gives a particular structure to the network that is generally not present in a MANET.

The network density and mobility model of a VANET can also vary significantly both temporally and geographically, as it follows directly from road traffic density. At peak hour in a metropolitan area, nodes will be closely packed together and moving slowly, but on a rural highway at midnight, nodes will be sparse and moving rapidly. Depending on node density, a VANET can also be very large — potentially the size of an entire road network — though frequent partitioning of the network is also common and thus there may often be very small networks, or even isolated nodes.

When compared with other mobile ad-hoc networks, such as wireless sensor networks, VANETs also differ in their application requirements. Firstly, energy constraints are not as important as the on-board radio unit can draw on power from the vehicle and/or can be recharged regularly, as opposed to some wireless sensor networks which rely on battery or solar power and may be deployed in remote or inaccessible areas where node maintenance is difficult or impossible. This in turn means that each node can be made capable of greater computational and transmission power without compromising the lifetime of the nodes.

VANETs may also be able to make use of roadside base stations: fixed nodes that participate in the network and are typically connected to a backhaul network for connectivity to the wider Internet. While other MANETs may also make use of fixed base stations, for any given application it will typically be known in advance whether a base station will be available, or failing that, at runtime for the duration of the application's execution. However, due to the high degree of mobility in VANETs and the large geographical areas they can cover, base stations may be available only intermittently and for a brief period of time. This means that their use and role in the network will differ for a VANET as opposed to other MANETs.

Additionally, VANET applications will often have a different information model, particularly when contrasted with wireless sensor networks. There is usually no concept of specific source or sink nodes; rather information is directed to and from geographical regions. For instance, a collision warning is relevant only to the vehicles in the immediate area, whereas traffic information may need to be transmitted to any vehicles travelling on feeder links. Each node in the network is thus both a source and a sink for information, depending on the type of information and the location of the node in the network. The main nodes that are differentiated in the network, again depending on the application, are roadside base stations. Since they may be the only nodes connected to the wider Internet (and thus to traffic authorities), they can be both an important source of information and/or instructions which vehicles cannot otherwise obtain, and collection points for information about traffic conditions, regulation violations, or other data.

Existing and potential applications for VANETs inform requirements for lower-level functions of the network. Different applications have differing levels of requirements for timing and reliability of data delivery. For instance, a warning about an approaching emergency vehicle or an imminent collision needs to be delivered in a timely manner and with certainty, whereas for traffic information, delivery might be important but delay-tolerant, and for entertainment applications such as web browsing for passengers a delivery failure is not critical.

Routing and forwarding protocol choice and design is also affected by intended applications [60, 85]. Many applications for VANETs are geographical in nature: a collision warning should go out to vehicles in the immediate vicinity, information on traffic conditions should be sent to vehicles on feeder links, a notification about an emergency vehicle's presence should be sent to vehicles along its route, and so on. Aggregation of data is also common — a vehicle may be interested in the average speed on a road segment, or a traffic authority may want information about the traffic density across the network. In fact, it is relatively rare that traditional unicast routing, in which information is sent to a specific vehicle regardless of its position, is needed, with one exception being entertainment applications in which passengers have requested specific content.

The geographical nature of information in a VANET has two major implications: first, that routing protocols should also be geographical, with variants on position-based routing being commonly proposed, and secondly, that localisation — that is, knowing the positions of nodes in the network — is an important service required in VANETs. Localisation in VANETs presents unique challenges due to the highly mobile nature of the network, however, this can be overcome through the use of on-board sensors (which might include GPS receivers or laser rangefinders) and there has been much work undertaken in this area.

3.4.1 Channel Characteristics

VANETs share many channel characteristics with other wireless networks. Like any wireless network, data is transmitted over radio waves, which are inherently a broadcast medium. This means that interference can occur if two nodes try to transmit at the same time, resulting in neither transmission being intelligible. In a network in which nodes are not all in transmission range of each other, this interference can even occur between nodes that are "hidden" from each other (i.e. not in range of each other), so long as there is at least one other node in range of both. This third node may receive transmissions from both simultaneously, resulting in interference and failed transmissions, but in this case the offending nodes cannot even sense that they have caused interference and cut short their transmissions, thus freeing the medium for use by other nodes. In VANETs, this hidden terminal problem is greatest when vehicles have high velocity [86], making it particularly problematic since this is also where message reception in a timely manner is most critical.

Another source of wireless reception problems is propagation loss. This is the attenuation of the signal as it travels over distance. Signals are stronger, and thus clearer, closer to the source than they are further away. This means that transmission range in a wireless network is not really a discrete concept. A signal will become progressively harder to receive the further away the sender is, until eventually it cannot be discerned at all. Moreover, because of other sources of signal attenuation, this distance is not uniform and thus cannot be easily predicted in advance and may change over time. This means that a node that is in range for one transmission may be out of range for the next or vice versa, particularly in a highly mobile network such as a VANET.

Radio waves, like other waves, can also have problems with fading, that is, variation in the attenuation of the signal, sometimes to the point where the data can no longer be understood. Fading can result from a number of causes. Shadowing occurs when an object that is not permeable to the waves lies in between the sender and receiver, blocking the transmission and preventing the sender from receiving the information, just as a building might block the sun resulting in a shadow on the ground. Multipath fading occurs when the same signal takes multiple paths to arrive at a destination, potentially resulting in destructive interference when it gets there — this would be similar to seeing the same light directly and in a mirror at the same time, however, the difference is that with a data transmission, it is not enough merely to detect the signal, its form must be sufficiently intact for the data to be decoded. Another issue can be the Doppler effect, in which successive wavefronts become either closer together or further apart, distorting the signal, due to the relative movement of the sender and receiver.

All of these types of fading are particular problems for VANETs even more so than for other wireless networks. In a VANET, vehicles themselves — being large physical objects, particularly certain types of vehicles such as trucks and buses — can cause shadowing or multipath fading, as can built-up urban environments. The high relative speeds of vehicles can also result in Doppler effects. Moreover, VANETs present their own unique challenges over other wireless networks, such as varying node density. A traffic jam can result in many nodes packed tightly into one area, causing a high level of contention in the network — since it is a broadcast medium, only one node may transmit at once — resulting in a lower data rate per node. For a more detailed explanation of VANET channel models, see [87].

3.4.2 Standards and Protocols

We present here an overview of the existing standards for vehicle-to-vehicle communication. Transitions from one standard to another are difficult and time-consuming in the context of vehicles, as a new standard must first be agreed to by all parties, including automobile manufacturers and traffic authorities, then tested thoroughly for safety, then finally rolled out to all vehicles — a lengthy process given that the lifespan of a vehicle can be decades. This means that the standards currently being developed and deployed are likely to remain widely used for some time. It should be noted that the work presented in this thesis is intended to be generic to any vehicular ad-hoc network and thus is not tied to any particular standard.

The main standard that is currently used for VANETs is IEEE 802.11p [88], an amendment to the IEEE 802.11 standard for use in vehicular environments. This standard covers the physical and MAC layers of the network. For higher-layer protocols, standards used are IEEE 1609 [89] in the US and ETSI EN 302 665 [90] in Europe. Among other functions, these latter two standards provide for periodic beaconing of messages containing a range of useful information about the current state of each vehicle (Cooperative Awareness Messages (CAMs) and Decentralized Environmental Notification Messages (DENMs) for ETSI and Basic Safety Messages (BSMs) for IEEE). Some examples of information contained in these messages are vehicle location (longitude and latitude), speed, acceleration and heading. This means that when developing applications for VANETs, if the presence of these standards is assumed, some or all of the information required by the application may be already available at no extra cost in terms of bandwidth usage or processing time.

In the US, 75 MHz of spectrum in the 5.9 GHz range has been allocated for dedicated short-range communications (DSRC) channels for use by intelligent transportation systems. In the EU, 30 MHz has been allocated, to be expanded in the future. Each channel in IEEE 802.11p uses 10 MHz, so this allocation allows for multiple channels, potentially with different purposes. In particular, safety-critical applications have a dedicated channel in both jurisdictions.

3.4.3 VANET Applications

There have been many applications proposed and implemented for vehicular ad-hoc networks. Here, we focus on those that have potential impacts on road safety, efficiency, or both. Some types of VANET applications and examples of each are listed below.

- General (non-safety) information dissemination [91, 92], in particular for information about traffic conditions and congestion [93–95]
- Safety-critical information dissemination, in particular
  - Incident notification [96, 97]

    - Dangerous surface condition warning [96]
    - Collision warning [56–59]

- Emergency vehicle warning systems [98]
- Platooning (i.e. groups of vehicles with co-ordinated longitudinal control) [96, 99–108]
- Intelligent merging [96, 109, 110]
- Intersection management [111–113]

Different ways of classifying these applications have been proposed in the literature. [60] uses an approach focusing on the purpose of the applications, with applications divided into categories of public safety, traffic management, traffic co-ordination and assistance applications (such as platooning), and traveller information support. A more recent survey [114] focuses instead on the communications requirements of applications. Using this approach, applications are separated into general information services, for which data is not time-critical and lost data does not have safety implications, vehicle safety information services, requiring time-critical and reliable message delivery to vehicles in a certain area, individual motion control, which is again time-critical but constrained to nearby vehicles and group motion control, in which messages must reach all the vehicles in a specific group with low latency.

The applications listed above focus either on a small-scale situation — an individual vehicle, a small group of vehicles in a platoon, or a single intersection — or else are systems designed only for disseminating information, rather than also controlling vehicles or managing traffic. This latter type thus do not form a closed control loop; although it is assumed a driver will respond to the information provided, this is generally considered external to the system itself and not modelled. The work described in this thesis differs from these previous applications in that it concerns a large-scale traffic management system, aiming to provide statistical improvements across the entire system by managing risk levels and thus accident rates and by modifying vehicle behaviour accordingly.

However, such a system could be used in conjunction with these previous applications. Co-operative driving systems such as those listed above add to the possible behaviours that vehicles can employ in response to risk levels. An understanding of risk can thus be used to inform the behaviour of both individual vehicles and groups of vehicles, helping to determine which of these co-operative behaviours should be taken at any given time or the exact parameters to be used for them.

3.5 Simulation Tools

The two main software packages used in conducting my work were Quadstone Paramics [115], for road traffic simulation, and ns-3 [116], for simulation of wireless networks. For some experiments, these tools were used together to provide a complete simulation of a VANET. Each of these programs is described in more detail below.

3.5.1 Quadstone Paramics

Paramics originated as a University of Edinburgh project in the early 1990's before being commercialised. It has been in continual development since then and is now in its sixth version. Paramics has been used by thousands of organisations in over 45 countries [115]. It includes a wide range of features for modelling road networks and vehicles as well as running and analysing simulations. Simulations can be run across multiple machines in parallel (without the need for additional licenses).

The main features of Paramics which make it suitable for the research in this thesis are its microsimulation model and programming API. The car-following model implemented in Paramics is based on the one discussed in [83]. Evaluations of this model and of Paramics' traffic simulation in general can be found in [117] and [118]. The use of a microsimulation model means that each driver and vehicle is modelled as an individual agent, with its own stimulus inputs and behaviour, and these agents are independently controlled in a parallel processing fashion each time step. Since we model behaviour of individual vehicles and the interactions between them, such a microsimulation model is not only suitable but even necessary.

Paramics is also highly programmable. Various aspects of the internal model can be overridden, or the model can be extended in a variety of ways. It is also possible to programmatically retrieve information from the simulation as it is running and set a wide range of parameters in the system. This means that it has been possible to implement new models and integrate them with the existing Paramics ones, as well as collect data for experimental results.

3.5.2 ns-3

ns-3 [116] is a discrete-event network simulator, that is, it does not have a universal simulation timer but instead schedules events, each with a time associated with them, and then executes these events in chronological order. It is licenced under the GNU GPLv2 licence and is freely available. Since ns-3 is open-source, any aspect of its models or operation can be modified by the user, although for our work this was largely unnecessary and only an application for ns-3 needed to be developed.

ns-3 contains a variety of networking models available for use; in particular we have used the `YansWifi` physical layer model [119] and the `NqosWifi` model for the MAC layer, as these provide the closest available fit to the 802.11p model. While there is some work in development on implementing full 802.11p models for ns-3, the differences to the 802.11 standard model do not impact the experiments carried out in this work.