EXPLORING THE APPLICATIONS OF CONNECTED VEHICLE DATA TO REAL-TIME SAFETY OPTIMIZATION AT ISOLATED INTERSECTONS

by

Tarek Ghoul

B.ASc., The University of British Columbia, 2019

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Civil Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2021

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, a thesis entitled:

Exploring the applications of connected vehicle data to real time safety optimization at isolated intersections

submitted by Tarek Ghoul in partial fulfillment of the requirements for

the degree of Master of Applied Science

in Civil Engineering

Examining Committee:

Dr. Tarek Sayed, Professor, Department of Civil Engineering, UBC
Supervisor

Dr. Paul de Leur, Manager of Road Safety Programs at Insurance Corporation of British Columbia
Supervisory Committee Member

Abstract

The proliferation of Connected Vehicles and their ability to collect a large amount of data present an opportunity for real-time safety optimization of traffic networks. At intersections, Adaptive Traffic Signal Control (ATSC) systems and dynamic speed advisories are among the proactive real-time safety interventions that can assist in preventing rear-end collisions. This thesis proposes two systems that utilize connected vehicle data to optimize traffic safety in real-time using reinforcement learning approaches. The first system utilizes a Deep Deterministic Policy Gradient (DDPG) reinforcement learning agent in conjunction with a dynamic programming approach to optimize vehicle trajectories and issue speed advisories. The second proposed system is a Signal-Vehicle Coupled Control (SVCC) system incorporating ATSC and speed advisories to optimize safety in real-time. By applying a rule-based approach in conjunction with a Soft-Actor Critic Reinforcement Learning framework, the system assigns speed advisories to platoons of vehicles on each approach and extends the current signal time accordingly. Dynamic traffic parameters are collected in real time and used to estimate the current conflict rate at the intersection, which is then processed and input into the respective models. The systems were tested on two different intersections modelled using real-world data through the simulation platform VISSIM. Significant reductions in traffic conflicts and delay were observed, with the simple speed advisory system yielding a 9-23% reduction in traffic conflicts and the SVCC system yielding a 41-55% reduction. Similarly, delay reductions of about 24% were observed. Both systems function at lower levels of market penetration, with diminishing returns beyond 50% Market Penetration Ratio (MPR). The thesis thus proposes and demonstrates the effectiveness of two unique CV-based systems that are low in computational intensity and applicable in the near future.

Lay Summary

Recent developments in intelligent transportation systems have allowed for new traffic safety techniques to be developed. Connected Vehicles are rapidly being adopted and their ability to transmit and receive data may be used to significantly improve road safety. Using these data, the real-time safety on a given span of roadway may be quantified, optimized and thus, improved. This thesis explores two real-time safety interventions. The first proposed system uses a reinforcement learning approach to provide individual vehicles with speed advisories to optimize for net safety. The second proposed system combines speed advisories with traffic signal extension to create a reinforcement-learning-based system that optimizes for safety by modifying traffic signal timing and issuing speed advisories. These systems were developed such that they are applicable to real-time environments with low computational intensity and tested in simulation environments with mixed traffic conditions.

Preface

This dissertation is the original intellectual product of the author, T. Ghoul under the guidance and support of Dr. T. Sayed. Portions of this thesis have been previously published in journals, or are set to be published as follows:

1) Portions of the literature review in Chapter 2 and portions of Chapter 3 and 5 have been published [Ghoul, T and Sayed, T. 2021 "Real-Time Safety Optimization of Connected Vehicle Trajectories Using Reinforcement Learning." Sensors, 21]. I was responsible for conducting most of the conceptualization, analysis, and writing. Dr. Sayed T. was the supervising author and was involved in concept formation and manuscript edits.
2) Portions of the literature review in Chapter 2 and portions of Chapter 4 and 5 have been submitted for publishing [Ghoul, T and Sayed, T. 2021 " Real-Time Signal-Vehicle Coupled Control: An Application of Connected Vehicle Data to Improve Intersection Safety" Accident Analysis and Prevention]. I was responsible for conducting most of the conceptualization, analysis, and writing. Dr. Sayed T. was the supervising author and was involved in concept formation and manuscript edits.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I would like to thank Dr. Tarek Sayed for his guidance and assistance throughout my academic career at UBC. He has been an excellent mentor who not only guided me throughout my time as a Master’s student, but encouraged me to continue and pursue academia. Between his instruction in the classroom and his enthusiasm during the research process, I am honored to have had him as my supervisor. Without him, this paper would not be possible.

I would also like to thank Mohamed Essa for his critical support and assistance during the early stages of this research, and for laying the foundations for this work. I would also like to thank my colleagues at the research team in the transportation research group, Mohamed Bayoumi and Ahmed Bayoumi, for their assistance in the lab. Furthermore, I would also like to thank my instructors during my time as a Master's student at UBC, Dr. Russell, and Dr. Bigazzi, for their assistance in the classroom and their understanding during a challenging period.

Lastly, I would like to thank my sister and my parents who encouraged me and supported me in countless ways during this process and beyond

Dedication

To my parents

Chapter 1: Introduction

1.1 Challenges

Most modern transportation networks have been designed for private vehicles. These vehicles can travel at high speeds and can move both individuals and goods. Unfortunately, the presence of human error in the system can result in frequent collisions. When coupled with high-speed complex systems, severe collisions are almost inevitable. While some policymakers have attempted to shift towards a more multi-modal transportation system, private vehicles will likely remain prominent for several years to come. As a result, measures should be devised to work towards preventing as many collisions as possible, or at the very least, minimize their severity.

Traffic collisions are among the leading causes of death worldwide. Policymakers have long attempted to minimize the damage caused by traffic collisions, which contributed to over $242B in losses in the United States in 2010 alone (Blincoe et al., 2015). In Canada, a nation with a relatively low collision rate, traffic collisions are the third leading cause of death and may be considered an epidemic (Statistics Canada, 2020). Many road jurisdictions have attempted to remedy the issue by establishing road safety interventions to reduce the number of collisions with varying levels of success. Several initiatives have also been established looking towards reducing the number of roadway fatalities to zero, such as the Vision Zero initiative. These initiatives have seen mixed results and further research is required to devise effective solutions to further mitigate the harm caused by vehicle collisions.

Significant work has been performed in the field of transportation safety attempting to devise traffic safety models that accurately identify regions or locations with significant collision risk. Historically, traffic safety models, which attempted to quantify the collision risk at road locations, utilized historical collision records to estimate the risk of future collisions. This

presented a moral dilemma as a statistically significant number of collisions (including fatalities and/or injuries) would have had to occur prior to any action being taken. Thus, a growing trend in safety research is the use of surrogate safety measures that do not rely on historical collision data such as traffic conflicts. Furthermore, it would be beneficial to devise a system that could simultaneously predict crash risk using surrogate safety measures (e.g., traffic conflicts) and subsequently develop intervenions that may reduce crash risk. Challenges concerning devising road safety interventions can broadly be classified into cost-related issues, data-related issues, and equity.

Cost-related challenges can potentially prevent the undertaking of a road safety intervention. Given that significant taxpayer contributions are required to finance and maintain road safety interventions, any measure will likely be extremely scrutinized by public authorities. While traffic collisions represent a significant economic loss to society, political will is not always available to implement costly remediation measures in problematic sites, even if they demonstrate a high potential for collision reduction. As such, cost constraints often prevent the implementation of novel systems with a high initial capital cost or systems with high maintenance. There is thus a preference for the implementation of several smaller and less effective, but proven interventions when possible. As a result, new road safety intervention proposals should not require a large amount of new infrastructure and should have a low initial capital cost with demonstrated reliability.

Data-related issues present another significant challenge. Data collection correlates highly with investment in traffic safety programs. In many cases, particularly in the developing world, the lack of consistent and reliable traffic safety data is a notable obstacle that prevents decision-makers from identifying the proper course of action. To propose effective solutions, the available

data must be of adequate quality and contain a statistically significant number of data points. In many cases, the lack of traffic count or collision data for specific localities presents a significant barrier in determining an adequate solution.

Although road safety interventions may be effective at improving safety and usually, mobility, the installation of the necessary infrastructure may be costly. There is a clear trade-off between data quality/quantity and cost. This poses an equity issue as there are limited resources available to implement novel solutions. The situation is complicated by the fact that certain interventions may improve one group's driving experience while adversely affecting another's. For example, solutions relating to traffic signal timing may be effective in improving the situation for one group of vehicles, or approach, but may adversely affect the performance of the system elsewhere. Hence, a holistic and low-cost approach is favorable.

In addition to the challenges associated with data availability and cost when implementing current "state of the practice" road safety interventions, the implementation of real-time solutions poses an even greater challenge. In real-time safety optimization, both computational power and available data are valuable assets. In many cases, it may be difficult to obtain relatively aggregate measures such as traffic volumes, turn movements, and collisions. Real-time data is even more difficult to acquire. For applications at intersections, such as speed advisory systems or adaptive traffic signal control, more granular data is required. Individual vehicle trajectories as well as kinematic variables such as speed and acceleration are usually required to develop effective systems. Traditionally, these systems have depended on either static data collection systems or computer vision techniques. Using data from these sensors, it would be possible to devise real-time safety optimization systems which rely on either modifying individual vehicle behavior or modifying the signal timing plan. Individual behavior may be altered by issuing advisories to

drivers in the form of a roadway sign, or by communicating directly with vehicles that have communication capabilities. Signal timing can be modified using a wide variety of adaptive traffic signal control (ATSC) algorithms that have previously been developed.

New technologies have been devised to address the problem of implementation cost at a large scale, and human error. These technologies have the potential to disrupt the field of transportation. Much like the development of the first automobile and the subsequent construction of large highways, the development of new intelligent transportation systems (ITS) may have a similar impact on modern transportation networks. Many of these emergent technologies have yet to be fully tested. Consequently, it is important to apply these new technologies to the field of traffic safety and ensure that the theoretical solutons are transferable to the real world.

1.2 Motivation

This thesis is predominately motivated by the rapid development and deployment of new transportation infrastructure technologies. By studying and applying the projected benefits of these systems, the incidence of collisions may potentially be greatly reduced. This thesis considers several innovations and relates them to the field of traffic safety through a unified framework. The following sections provide additional details about the motivations of this study.

1.2.1 Intersection Safety

Intersections pose a significant risk to collisions, as over 45% of collisions in Vancouver occur at intersections, many of which are rear-end collisions (ICBC, 2020). This represents considerable economic losses that may be preventable. Most of these collisions typically involve some degree of road user error, exacerbated by unstable traffic flow and other site specific issues in collision prone locations. It is thus a moral obligation to prevent and reduce the incidence of these types of collisions if possible. Real-time strategies are effective and may be employed for

these purposes. In line with established traffic flow theory, the conflict rate and collision risk at an intersection may be estimated and reduced by either altering the trajectory of individual vehicles or by modifying the signal timing plan. This thesis will focus on applying these concepts to intersection environments.

1.2.2 Traffic Conflict Estimation Techniques

To improve the safety at intersections in real-time, an effective real-time safety performance metric should be used. Precedent studies used readily measurable quantities, obtained from sensor data, to develop safety performance functions that can quantify safety. These metrics were designed at the phase level, but can be applied to smaller timeframes. This would allow for both real-time monitoring of safety at intersections, as well as real-time interventions to reduce traffic conflicts. As such, this thesis seeks to utilize newly developed safety performance formulas to propose systems that may improve intersection safety.

1.2.3 Projected State of Technological Advancement

Connected vehicles (CV) are among the technologies that are rapidly proliferating in recent years. The market share of CVs has drastically increased, with multiple projections estimating that the majority of vehicles produced by 2045 will have CV capabilities (U.S Department of Transportation, 2015). CVs present an opportunity to optimize road networks for safety and mobility in real-time. By utilizing vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), and infrastructure-to-vehicle (I2V) communication systems, connected vehicles provide situational awareness to both the infrastructure that they utilize, and to the vehicle users themselves.

CVs are complemented by other technologies and advancements in processing power and connectivity which allow for large amounts of data to be collected and processed in real-time. The phenomenon of “Big Data” when combined with CVs will dramatically alter the roadway,

presenting an opportunity to develop novel techniques to reduce traffic conflicts. This thesis will explore potential applications of CVs using both V2I and I2V communication as they apply to safety.

1.2.4 Connected Vehicle Applications

Several CV systems intended to help improve safety and throughput have been devised using both V2I as well as I2V communication systems. Green Light Optimal Speed Advisory (GLOSA) systems are an example of such systems. GLOSA allows for the optimization of vehicle trajectories by issuing speed advisories to connected vehicles, indicating whether to speed up or slow down. Large improvements in safety and mobility are expected as a result of these systems. Unlike autonomous vehicles, connected vehicles are not actuated by a central controller, and speed advisories are issued as recommendations to a human driver. As such, compliance is a key consideration in any proposed GLOSA-based framework.

Adaptive Traffic Signal Control (ATSC) is another promising application of CVs. By receiving data from discrete vehicles, a central signal controller can adjust the signal timing at an intersection to improve safety and mobility. Traditional ATSC systems rely primarily on fixed infrastructure for data or computer vision techniques. These are typically costly to implement on a network level. With a large CV fleet, data may be available at every intersection and as such, it would be easier to optimize traffic signals dynamically rather than rely on long-term data for signal timing.

Another interesting application involves combining ATSC systems and GLOSA to jointly control vehicle trajectories and signal timing. Signal Vehicle Coupled Control (SVCC) would amplify the benefits of each system and would issue speed advisories to CVs based on a dynamically changing signal timing plan. This thesis will explore the applications of CVs at

intersections as they relate to trajectory optimization systems alone, as well as joint Signal-Vehicle Coupled Control systems.

1.3 Problem Statement

With the expanded use of CVs, several opportunities begin to present themselves with respect to improveing intersection safety. Systems with limited applications that have traditionally relied on fixed sensors may now utilize a large amount of data to further improve network performance. Unlike existing systems which rely on fixed sensors, CVs have the potential to provide a low-cost and high-efficiency approach to improving intersection safety by making real time data more accessible. Further enhancements may be observed should vehicle trajectories also be optimized through speed advisories from intersection control systems. This thesis seeks to explore different applications of CVs in relation to intersection safety.

1.3.1 Problem One: Real-Time Trajectory Optimization Using Reinforcement Learning to Improve Intersection Safety at Isolated Intersections

The optimization of safety at intersections is a difficult feat that poses several challenges that must be overcome. Traditional optimization systems rely primarily on aggregate data and signal timing plans in conjunction with conditional logic to work towards minimizing stop time at intersections. By establishing a situation with fixed traffic signals and utilizing prior data, the system can guide a vehicle through speed advisories such that it arrives during a “green wave”. Alternatively, signal controllers can modify signal timing to accommodate additional vehicles based on sensor occupancy. More sophisticated theoretical models have attempted to use linear optimization techniques to accomplish these tasks and have demonstrated large improvements in safety and mobility. While effective, many of these systems lack transferability to real-world applications as linear optimization techniques may take a long time to process. The development

of new reinforcement learning (RL) techniques has allowed for the creation of new, computationally simple algorithms that can be used for optimization. With a computational complexity similar to that of a regression model, these systems have great potential in the field of safety optimization if the appropriate data are available.

Vehicle trajectory optimization is a potentially useful application of connected vehicles and real-time data analysis. Using speed advisories, it may be possible to reduce the incidence of collisions by warning vehicles ahead of time of upcoming phase changes and advising them to either speed up or slow down accordingly. Despite increased speed often being detrimental to safety, reducing the number of vehicles present at the intersection during a red light, under the appropriate conditions, can result in reduced incidence of rear end collisions caused by stop and go waves. While systems such as GLOSA have already been tested, numerous problems must be addressed when considering an isolated intersection with a highly connected fleet. Firstly, many precedent GLOSA systems operate within a framework of multiple intersections or long road spans, assuming that there are either sensors located along the entire span of the roadway or that said sensors connect directly to the vehicles through an internet connection. While this assumption may be valid, it would be useful to develop a GLOSA system that would work in shorter roadway spans, given that the average direct short-range communication sensor (DSRC) has a range between 150 m and 300 m, or an average of 225 m (U.S Department of Transportation, 2015). If placed at an intersection, there would be a short timeframe for vehicles to be detected, processed, and for speed advisories to be issued.

A further complication of the problem stems from the mechanism by which vehicle trajectory optimization techniques at intersections attempt to improve mobility and safety. These systems tend to work towards improving the arrival pattern over a span of roadway or several

intersections, thereby organizing vehicles into platoons that can cross an intersection at an ideal time. If a shorter roadway span is considered, then there will be less time for vehicles to be organized into platoons effectively. The problem becomes more complicated when considering multi-lane approaches. Further, given the real risk of non-compliance if implemented in practice, any given solution should observe the effects of non-compliance on the system through modeling. Thus, this establishes the following research goal:

Develop a real-time trajectory optimization algorithm that can be used at an isolated intersection to reduce traffic conflicts while only considering and issuing speed advisories to CVs within the DSRC range. The system must be functional for intersections with multi-lane approaches and should consider the effects on Platoon Ratio as well as the potential issue of non-compliance.

1.3.2 Problem Two: Incorporating Adaptive Traffic Signal Control using Connected Vehicle Data

Prior studies have considered the application of connected vehicles as they apply to adaptive traffic signal control systems. By providing real-time data regarding individual vehicle positions, a traffic signal controller may alter the signal timing plan at the intersection. In turn, this reduces stopped time which is conducive towards reducing rear-end collisions. Vehicle trajectories can be estimated once transmitted to a central controller, thus providing the controller with additional data for safety optimization. By applying a framework similar to that discussed in previous sections, it may be possible to develop a new RL-Based algorithm that can optimize for safety in real-time.

Given that CVs may transmit and receive data, vehicle-side and intersection-side systems may be used to help reduce traffic conflicts at intersections. ATSC systems as well as GLOSA

systems have been demonstrated to greatly reduce the incidence of rear-end traffic conflicts by minimizing stopped times. As such, it may be possible to amplify the safety benefits of each system by using a CV-based ATSC system which improves the arrival pattern at the intersection through trajectory optimization as well. This is known as Signal-Vehicle-Coupled-Control (SVCC). This presents the following research objective:

Develop a joint Signal-Vehicle Coupled Control system in which connected vehicles transmit data to a central controller which modifies the existing signal timing plan and issues speed advisories to connected vehicles accordingly. The system must be functional in the transitional state where not all vehicles are CVs and should be applicable in real-time.

1.4 Contributions

This thesis seeks to explore safety-related interventions at intersections using CV data and presents two algorithms that may work towards achieving this goal. Both algorithms have been tested in simulation environments and were developed using real-world data and with the intention of being applied to the real world.

1.4.1 Real-Time Single-Agent Reinforcement Learning Based Trajectory Optimization Algorithm for Isolated Intersections

The first algorithm presented in this thesis consists of a mixed DDPG-dynamic programming algorithm that optimizes vehicle trajectories at intersections. Using a single agent to optimize the trajectory of CVs on four separate multi-lane approaches simultaneously, the system demonstrates that short-range trajectory optimization is possible. Furthermore, the single-agent system establishes the possibility of application to ATSC systems in the future as part of a collaborative system.

1.4.2 Real-Time Signal-Vehicle Coupled Control Algorithm

The second algorithm which is presented in this thesis consists of a mixed Soft-Actor Critic and dynamic programming algorithm which optimizes both vehicle trajectories and signal timing simultaneously. Unlike prior SVCC systems, this system is functional in real-time and utilizes CV data to quantify the current conflict rate and utilize it as an input. The algorithm was tested on numerous scenarios and appears to be both transferable and functional in the transitional state.

1.5 Thesis Structure

This thesis is organized into five chapters. The following breakdown details the contents of each chapter:

Chapter 1 introduces the thesis by exploring the state of traffic safety as it relates to connected vehicles and detailing the motivations behind this study. The section concludes by detailing the research objectives and listing the contributions of this work.

Chapter 2 provides a background of previous studies performed relating to safety performance functions, reinforcement learning algorithms, trajectory optimization systems, adaptive traffic signal control, and signal-vehicle-coupled control.

Chapter 3 explores the first study which details the development of a hybrid DDPG/dynamic programming trajectory optimization algorithm for isolated intersections.

Chapter 4 explores the second study which details the development of a hybrid SAC/dynamic programming Signal-Vehicle Coupled Control system.

Chapter 5 concludes and draws upon the contributions of prior sections and the lessons learned, paving the way for future work in the field.

Chapter 2: Literature Review

This section reviews the available literature as it relates to data extraction using CVs, trajectory optimization, adaptive traffic signal control systems, reinforcement learning algorithms, as well as relevant vehicle behavior that relates to the systems proposed in this thesis.

2.1 Traffic Information Extraction

The performance of any fully developed model depends on the quality and quantity of data that can be obtained. As such, readily available CV data measured with reasonable accuracy should be used as inputs for the model. Accurate sensor data is necessary for the development of an algorithm that can both approximate the current state of the environment and improve upon it in a way that is applicable to real-world conditions.

Several studies have been performed using V2I and V2V communications to extract traffic information. A review of the literature indicates that basic parameters such as vehicle positions and queue length are readily available and can be obtained in practice. Previous studies have utilized multiple data sources to extract traffic state data from an environment including the use of fixed detectors positioned on two points (Hao et al., 2013, 2015), probe vehicles with V2I capabilities (Hiribarren G, 2014; Hofleitner et al., 2012), and on-board CV GPS data (Comert & Cetin, 2009, 2011). According to a review of CV data collection techniques, techniques to parse and analyze these traffic data are either shockwave-based, kinematic-based, or stochastic learning-based (Guo et al., 2019).

Applying a shockwave-based method, aggregate data from multiple vehicles or platoons of vehicles are used to derive vehicle trajectory curves and identify boundary conditions. The parameters derived are then used to extract key shockwave parameters (Y Cheng et al., 2013). While a degree of uncertainty may be expected using this method, this method demonstrates the

possibility of obtaining parameters such as queue length and shockwave area through V2V and V2I sensors. Alternatively, kinematic-equation-based approaches utilize discrete vehicle data and collect basic information about vehicle positions, speed, and acceleration (Hao et al., 2015). Since these approaches require many parameters, the accuracy of traffic state approximation would depend on the number of vehicles in the fleet. Certain stochastic and deep learning techniques seek to bypass this requirement by developing more sophisticated statistical techniques (Zheng, 2017).

Given the abundance of available techniques detailing numerous approaches for data collection and analysis, CVs can be expected to provide basic data about their positions and speeds, and by extension, shockwave parameters detailing the traffic state. This paper utilizes the kinematic and shockwave data derived from these techniques as inputs to the proposed model as the literature supports the assumption that they are readily available.

2.2 Real-Time Safety Evaluation

Real-time safety evaluation is necessary to develop any system or intervention to improve safety. Waiting for a statistically significant number of collisions to occur to identify safety issues and implement countermeasures presents an ethical dilemma. Relying upon long-term crash data results in significant damage to people and property which may not have occured had the situation been prevented earlier. Conflict data is commonly used as a surrogate for collisions to gauge traffic safety, with measures such as Time to Collision (TTC), Post Encroachment Time (PET), and Deceleration Rate to Avoid Crash (DRAC) being commonly used as performance measures (Hayward, 1972). Modified versions of TTC and DRAC have also been devised to approximate collisions using different definitions of traffic conflicts (Fazekas et al., 2017; Ozbay et al., 2008).

Several studies have attempted to develop Safety Performance Functions (SPFs) to relate dynamic traffic parameters to conflicts and collisions at the cycle level. Data relating to shockwave parameters such as shockwave area and queue length have been used to evaluate safety at the cycle

level, with the same measures generalized to provide real-time instantaneous measures (Essa & Sayed, 2018). These SPFs will be used to approximate the performance of the model used in this study.

2.3 Platoon Behavior

Platoons can be defined as a group of vehicles moving along a specific approach categorized by several predetermined parameters. Some studies classify vehicles as a platoon based on the headway of following and leading vehicles, (Kamonthep et al., 2019) while others utilize thresholds such as the number of vehicles within the group, density, or speed differentials to establish similar states and group vehicles together into a platoon (Baibing, 2017; Gaur & Mirchandani, 2001; Yingyan et al., 2016).

Platoons can be used strategically as part of signal optimization algorithms to improve safety and throughput (He et al., 2012). By controlling the formation and dissipation of incoming vehicle platoons, the arrival pattern at the signal stop-line can be altered to improve system performance. This is highly dependent on both the initial arrival pattern and the available time for platoon formation and dissipation to occur. ATSC systems such as PAMSCOD have been previously developed using a platoon-based optimization algorithm to demonstrate the value of platooning in an ATSC context (He et al., 2012).

The performance of a system using emergent technologies such as CVs and connected autonomous vehicles (CAVs) is challenged under conditions where conventional vehicles share the road space. While it is possible to detect these vehicles using alternative sensors, conventional vehicles would not be responsive to direct speed advisories as they lack an onboard information suite. The trajectory optimization of conventional vehicles may be performed by exploiting platooning behavior. Additionally, pacer vehicles leading a platoon may be used to regulate the speed of following conventional vehicles. (Behl & Mangharam, 2010). A study conducted by

Kreideh et al. in 2018 found that even with a less than 2.5% market penetration ratio (MPR), the use of autonomous vehicles in freeway environments greatly reduced the magnitude of stop and go waves, improved traffic stability and, improved throughput (Kreideh et al., 2018). Therefore, it can be inferred that a functional trajectory optimization algorithm for mixed traffic conditions can be developed provided a certain percentage of the platoon is well situated and can receive speed advisories from the signal controller.

Consiering individual vehicles in platoons, the presence of connected vehicles may alter both individual and group behavior. Connected vehicle behavior may be distinct from conventional vehicles concerning breaking and acceleration. This results in reduced time-to-collision during car-following and a lower deceleration rate to avoid crashing (Ali et al., 2020). Consequently, drivers in the connected environment also maintain a higher safety margin with respect to pedestrian crossings (Haque et al., 2021). This safety margin is also likely to translate well to intersections. Driving simulation studies have shown that connected vehicles at intersections are less likely to run yellow lights (Ali et al., 2021). The compound effect of this prudent driving behavior has yet to be tested in the field due to the lack of implementation. Consequently, existing software packages modeling conventional driving behavior may be subject to a degree of uncertainty. Moreover, the opposite is also likely to affect results. Conventional vehicle drivers may react poorly in mixed traffic conditions due to the difference in driving behaviour between them and connected vehicles. Additional field testing is required to observe these effects and model them with precision.

2.4 Previous Vehicle Trajectory Optimization Algorithms

2.4.1 Variable Speed Limits

The implementation of speed advisories for specific connected vehicles may result in behavior similar to that observed on roadways with Variable Speed Limits (VSLs). Unlike

connected autonomous vehicles (CAVs) and autonomous vehicles (AVs), CVs require a compliant human driver. Much like driver behavior around speed limits, there will likely be a degree of variance between the speed limit and the actual observed speeds. VSLs have been used since the 1960s to regulate traffic flow and improve throughput. By reducing the speed variance between upstream and downstream vehicles on highways, traffic instability is reduced, thus reducing collision frequency. VSLs can operate either by sending a signal to individual vehicles or by being displayed on a roadside sign. Depending on the traffic conditions, different speed limits may be imposed to work towards achieving a critical speed that typically corresponds to a maximum flow at that point in time (Lu & Shladover, 2014). Alternatively, variable speed limits may be used to issue speed reduction advisories when road conditions are unsafe.

While VSLs have demonstrated their effectiveness on long roadway spans such as highways and reduced collision incidents in merging movements (L. Zhang et al., 2017), shorter roadway spans may affect their effectiveness. Nonetheless, the modification of upstream speeds may reduce bottlenecks and contribute towards improving the arrival pattern at the intersection (Liu et al., 2015).

A major limitation with respect to VSLs and speed limits, in general, is the issue of compliance. This is examined in a case study in Prague, Czechia, which concluded that despite the imposition of a variable speed limit in a ring road, there was insufficient evidence to suggest that there was a significant difference in speed resulting from the imposition of VSLs. A larger variance in speed was observed due to compliant drivers moving onto the right lane, and non-compliant drivers driving faster due to the lane change (Michał & Ondřej, 2016). As a result, proper enforcement measures must be taken should VSLs be implemented. Additionally, it is important to ensure that there is not a significant speed differential between the various lanes as the increased

variance in speed would result in additional collisions. Furthermore, any devised algorithm must account for a significant proportion of non-compliant vehicles in the connected vehicle context, resulting in additional modeling being necessary. For the purposes of this study, a non-compliant vehicle may be treated as a conventional vehicle with respect to market penetration ratio modeling.

2.4.2 Trajectory Optimization at Intersections

Vehicle trajectory optimization has been the subject of numerous papers in recent years, with studies focusing primarily on the implementation of connected autonomous vehicles (CAV) and their potential to augment the benefits obtained by ATSC systems. Many studies have demonstrated the benefits of trajectory optimization and demonstrated reductions in carbon emissions and delay. The developed systems either function separately to complement the benefits provided by ATSCs (Jiang et al., 2017; Wan et al., 2016), or work directly with the signal controller to establish a reservation-based system to allow vehicles to clear an intersection in time (Dresner & Stone, 2004; Ma & Li, 2021).

Green Light Optimal Speed Advisory (GLOSA) systems are amongst the most commonly used trajectory optimization techniques with regard to CVs. Previous studies have developed GLOSA systems for multiple-segment systems considering routing, as well as discrete signal segment localities. Multiple-segment systems, according to the literature, are more likely to provide improved performance, given that the traffic conditions are flexible enough to allow for a wide range of speeds over a larger distance (Seredynski et al., 2013). Similar field studies relating to the length of roadway segments have shown that providing drivers with signal state data earlier yields improvements to reaction time and safety (Iglesias et al., 2008). As a result, it is favorable to provide advisories as early as possible.

Studies involving the performance of GLOSA systems have shown promising results with benefits to fuel efficiency and delay reduction in high-density environments if applied

appropriately (Suramardhana & Jeong, 2014; Suzuki & Marumo, 2018). GLOSA algorithms that were previously derived rely primarily on conditional logic based on either current vehicle position in designated stop and go zones (Suzuki & Marumo, 2018), or on trajectory prediction for discrete vehicles (Karoui et al., 2018; Katsaros et al., 2011; Suramardhana & Jeong, 2014). Certain more complex methods include the use of genetic algorithms or fuzzy rule-based systems to reduce congestion over a longer span of roadway. (X. Zhang et al., 2014) At the most basic level, GLOSA presents the concept of a time to intersection (TTI) and establishes a means by which a TTI threshold is calculated and a favorable slowdown speed is obtained.

2.5 Adaptive Traffic Signal Control Systems

Several studies investigated the real-time safety benefits of ATSC systems. ATSC systems implemented include the Sydney Coordinated Adaptive Traffic System (SCATS) and Split Cycle Offset Optimization Technique (SCOOT). These systems show notable improvements to safety (Fink et al., 2016). Machine learning-based approaches which utilize safety performance functions and shockwave parameters as inputs have also demonstrated great potential to further safety improvements (Essa & Sayed, 2020).

A wide variety of techniques have been used in the literature to develop ATSC systems. Dynamic programming, among other methods, is commonly used in the literature. Among other techniques, dynamic programming was used to optimize for a specific reward function, allocating green time along both corridors and individual intersections to reduce delay (Beak et al., 2017; Feng et al., 2015; Islam & Hajbabaie, 2017). It is worth noting that the objective functions in these systems differ, with some optimizing for delay reduction and others considering fuel consumption. Simple RL techniques such as Q-learning (Arel et al., 2010; Camponogara & Kraus, 2003; El-Tantawy et al., 2013; Lu Shoufeng et al., 2008; Salkham et al., 2008) and Deep-Q-Learning (El-Tantawy et al., 2013; Gu et al., 2020; Tan et al., 2020) have been applied to ATSC systems to

optimize for key parameters. While most studies optimize single intersections, multi-agent approaches have also been attempted (Arel et al., 2010; El-Tantawy et al., 2013). These RL ATSC methods are typically optimized for queue length, delay, and travel time, but rarely considered safety (Essa & Sayed, 2020).

An emerging field of research however deals with the optimization of traffic signals using data from connected vehicles, but also through issuing recommendations to these vehicles to speed up, stop, or slow down (W. Li & Ban, 2017; Pandit et al., 2013; Xie et al., 2011). These studies utilize connected vehicles in three ways with respect to ATSC:

- Connected Vehicles may receive information from intersection signals via I2V communications to achieve objectives such as fuel efficiency or travel time.
- Signal Controllers can receive information from Connected Vehicles relating to platoon size, queue length, speeds, and other parameters to optimize for safety, delay, and fuel efficiency.
- Signal Controllers may utilize data from connected vehicles to optimize both the signal timing and the trajectory of vehicles with respect to safety, delay, and fuel efficiency.

The use of trajectory optimization with signal optimization is referred to as a Signal-Vehicle Coupled Control system (SVCC). Studies have been performed on the topic but typically relate to only delay or vehicular emissions (Jiang et al., 2017; Z. Li et al., 2014; Raj et al., 2015; Xu et al., 2017). Furthermore, these studies rarely consider computational complexity and rely on techniques such as mixed-integer linear programming or dynamic programming approaches (Z. Li et al., 2014; Wang et al., 2020).

Few studies have been performed on safety and SVCC algorithms with respect to connected vehicles, especially in real-time. The existing studies focus primarily on minimizing

delay based on optimizing stop-and-go behavior (Ilgin Guler et al., 2014), optimizing intersections by altering trajectories in fully automated environments (Z. Li et al., 2014), and optimizing capacity (Sun et al., 2017). The benefits of vehicle automation/coordination when compounded with ATSC have also been investigated (Xu et al., 2017).

2.6 Reinforcement Learning Techniques

Reinforcement learning (RL) is a subset of machine learning which can be used to develop a system that interacts with a dynamic environment in real-time. The ideal behavior of a system can be defined within the RL framework and the technique can be used to apply actions to an environment to achieve the desired result. Unlike in supervised learning, the system does not require a human observer to validate and verify the result and instead relies upon a reward function to evaluate the effectiveness of previous actions and adjust the future policy.

The optimization process varies depending on the specific reinforcement learning algorithm used. The most basic form of reinforcement learning is known as the Q-learning approach by which a matrix of all possible states, actions, and the expected impact of those actions is collected. Simple Q-learning excels in situations with relatively low complexity, discrete actions, and discrete environmental state observations. The Q function computes the reward of a given set of actions by using an equation known as Bellman's equation. This allows for the Q-matrix to be filled over time as the agent explores more results of state-action pairs. This process is known as exploration. The agent's policy defines the overarching decision to select the proper action given a certain state and is refined through further exploration. While Q-learning is an effective reinforcement learning technique, it scales with greater numbers of environmental input variables and actions leading it to suffer from the "curse of dimensionality".

To remedy this issue, the Q-matrix can be replaced with a neural network that may approximate the state-action pairs using precedent observations and rewards to calibrate the neural

network during the exploration process. The Deep-Q-Network (DQN) approach thus utilizes a neural network as a replacement for the Q-matrix and applies a similar framework involving a policy to decide upon subsequent actions. DQNs are designed for situations with a continuous observation space and a discrete action space. Each input node represents a continuous observation obtained from the environment and each end node in the neural network represents a discrete action that should be taken given a certain set of inputs. Situations where multiple actions can be taken can be represented as a larger action space, where every possible combination of the different actions and their discrete values are organized into a number of nodes. This results in a similar combinatorics problem with regards to high complexity problems requiring multiple actions. (Mnih et al., 2013).

Deep Deterministic Policy Gradients (DDPG) (Lillicrap et al., 2015), Proximal Policy Optimization (PPO) (Schulman et al., 2017), and Soft Actor-Critic networks (SAC) (Haarnoja et al., 2018) have been developed to allow continuous observation and action space. While the various methods function in different ways, the stochastic approach of the PPO and SAC systems are variations of an actor-critic framework. Actor-critic systems feature an actor-network deciding upon an action and a critic network approximating the value of said action to adjust the policy accordingly. Exploration is handled using an epsilon greedy parameter which represents the percentage chance of taking a sub-optimum solution contrary to the policy to explore potentially useful solutions. The SAC framework refines upon this concept to create a system that balances between entropy and expected return to avoid converging on local maxima. This simplification greatly assists in problems involving multiple continuous action spaces and allows for more optimum policy to be determined.

Chapter 3: Application of RL to Trajectory Optimization Algorithm

This chapter describes the development of a new trajectory optimization technique for connected vehicles at isolated intersections. Using a DDPG algorithm, vehicle trajectories are optimized through the assignment of distinct vehicle speeds to platoons of vehicles on a given lane and approach. The system is described in detail, and the methodology used to evaluate the system and quantify the benefits is examined. The limitations are discussed at the end of this section.

3.1 Developing a Speed Advisory System

This thesis considers two signalized intersections with four approaches operating under normal conditions. The study intersections were created based on real intersections in Surrey, BC, and real-time video data was used to ensure that the model accurately represents reality. Utilizing key concepts from precedent work, a reinforcement learning approach was applied to the problem and simplified using dynamic programming. The algorithm used can be described as microscopic and uses basic kinematic relations in order to optimize for safety. Thus, a dynamic programming approach is used to identify situations that warrant the determination of an ideal speed. A reinforcement learning algorithm then computes and issues these ideal speeds using observation data from the environment to optimize a reward function that minimizes traffic conflicts.

3.1.1 Selecting a Function to Quantify Safety

Essa and Sayed, 2020 developed several safety performance functions that can be used to quantify vehicular safety in short time intervals which allow for traffic conflicts to be estimated at the cycle level in the ATSC context (Essa & Sayed, 2020). The safety performance functions relate dynamic traffic parameters such as shockwave characteristics, and cycle parameters to estimate traffic conflicts. The most significant parameter used was the shockwave area which is strongly

correlated with rear-end conflicts at intersections. The derived safety performance functions are as follows:

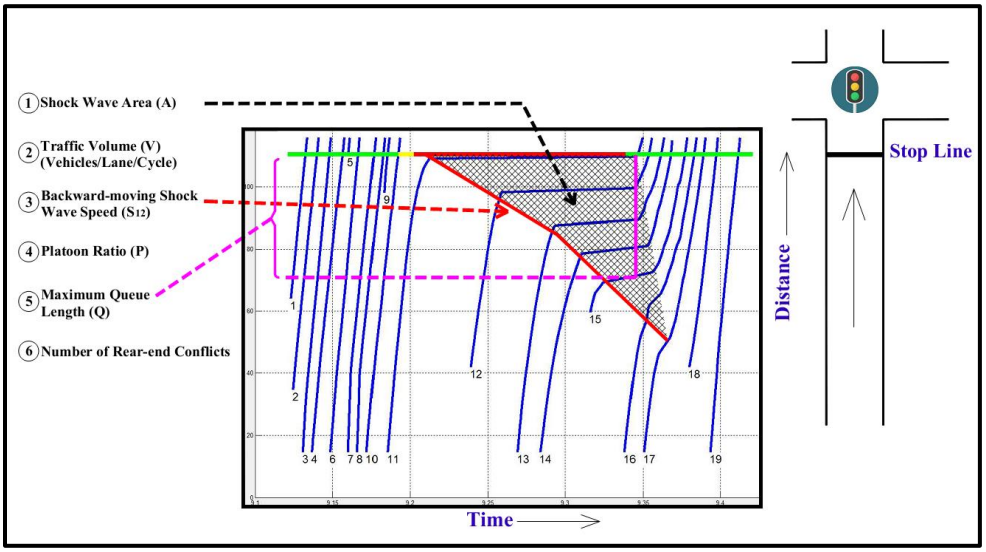

Figure 1: Shockwave Parameters and Traffic Conflict Models (Essa & Sayed 2018)

1. $\boldsymbol{E(Y)} = V^{1.563} exp(-3.231)$

2. $\boldsymbol{E(Y)} = V^{0.706} exp(-1.797 + 0.501\ A)$

3. $\boldsymbol{E(Y)} = V^{0.65} exp(-2.046 + 0.0122\ Q)$

4. $\boldsymbol{E(Y)} = V^{1.637} exp(-3.316 + 0.05\ S_{12})$

5. $\boldsymbol{E(Y)} = V^{1.571} exp(-1.768 - 1.266\ P)$

6. $\boldsymbol{E(Y)} = V^{1.239} \exp(-1.624 + 0.294\ A - 0.828\ P + 0.119\ S_{12})$

Given that this algorithm does not seek to optimize signal timing, signal level parameters such as platoon ratio may not be representative of the desired performance of the system. Ideal behavior would be characterized by vehicles slowing down to an ideal speed if they cannot clear an intersection or speeding up to clear an intersection, minimizing idle time at the intersection. Furthermore, for the purpose of real-time optimization, a measure that is easily obtained within

seconds is preferable to one which requires the entire cycle to process or relies upon rolling averages. Thus, the most useful parameters used in this study are shockwave area and traffic volume, corresponding to the second formula derived in the previous study. The shockwave area can be obtained at any given point in time. Changes in shockwave area may be instantaneous, making it a useful measure for real-time safety evaluations. Similarly, traffic volume is among the simplest variables to obtain, and changes are easy to observe.

Therefore, the model in Equation 2 was selected as the primary performance measure. Since the intersection which will apply these measures includes multiple approaches, the total safety at the intersection must consider all approaches. Thus, traffic safety can be quantified using the sum of expected conflicts for a given lane and approach as defined by Equation 2. With an appropriate measure of safety clearly defined, this will be used to improve vehicle safety at the intersection.

It should be noted that the functions derived in the previous studies only apply to rear-end conflicts. Given that these are the most common collisions at intersections, other conflict types such as sideswipes and left-turn conflicts were not considered and may be examined in future research

3.1.2 Platoon Recognition Algorithm

Given the large number of vehicles expected to be present in the system, organizing these vehicles into groups may be beneficial to simplify the problem. A review of the literature indicated that platoon identification is a contentious topic. Complex methods are used to identify clusters of vehicles in situations where more than one lane is present. In this thesis, a simplified version of the platoon recognition algorithms found in the literature was used.

As observed in the literature, some techniques identify platoons based on the minimum number of following vehicles, while others utilize a critical headway parameter to designate vehicles forming

part of a platoon (Kamonthep et al., 2019). After calibration, and through trial and error, a critical minimum platoon size of three vehicles, a maximum critical time headway of 5 seconds, and a space headway of 35 m was selected. These variables were selected and calibrated through testing and validated by visual inspection of the results using simulation software. The methodology used was based on the key variable of critical time headway identified by Kamonthep et al. 2019 (Kamonthep et al., 2019).

By utilizing this concept, platoons can be identified on single lanes. The framework works sufficiently when considering two separate adjacent lanes. During the calibration process, the observed likelihood of a cluster of vehicles on different lanes sharing the same longitudinal positions but being classified as two separate platoons was minimal. Nonetheless, it is necessary to account for this potential scenario should it arise. To modify this framework to adapt to multiple lanes, the first and last vehicles of the platoon on each lane are identified. Their positions are then transferred onto adjacent lanes. The maximum of this range, provided that the headways do not exceed the critical values, were used to further classify the same platoon across multiple lanes. Thus, the position of the first and last vehicle of the first platoon in Lane A was compared to the position of the first and last vehicle of the first platoon in Lane B, as well as adjacent platoons to recalibrate and develop a system that may be used for two-lane approaches.

This approach was primarily used to ensure that the model remained simple and easy to implement in real-time. Noting the inherent limitations of using a fixed critical parameter assumed via calibration, the platoon recognition algorithm used for the purposes of this thesis is simple and not as accurate as more complex methods. It would be possible to further refine this system by utilizing a supervised learning approach or by applying more complex pattern recognition algorithms. This system, however, remains accurate enough for the purposes of defining gaps

between identified platoons of vehicles and allowing the system to decide whether to close the gaps, create new gaps, and dynamically assign speeds to the various groups.

3.1.3 Visualizing the Ideal Individual Vehicle Behaviour

It is important to identify the ideal vehicle behavior such that a solution grounded in theory can be found. In theory, the ideal discrete vehicle behavior would be defined based on the signal state, with the vehicle accelerating or decelerating depending on whether the vehicle can safely cross the intersection. Should the vehicle decelerate to a low speed, there may be negative effects on traffic flow and potentially the creation of stop-and-go waves over time. Thus, in the event of a deceleration, there is a critical target speed that the algorithm should seek to obtain which would prevent additional shockwaves from being formed.

Furthermore, platooning behavior should be considered. Should a vehicle be asked to slow down, it would likely join with the preceding vehicles to form a platoon with the first leading vehicle regulating the speed of vehicles behind it. Vehicles asked to speed up may similarly join with vehicles ahead to increase the size of a given platoon. Individual vehicle behavior would compound into group behavior that would result in entire platoons forming, dissipating, splitting, or joining together.

Noting the integration between platoon group behavior and individual vehicle behavior, the ideal solution would ultimately depend on signal state and projected trajectory, allowing for a set of different solutions to be established and applied to entire platoons of vehicles.

3.1.3.1 Scenario A: Vehicle Arrives at Intersection During Green Phase

In the first scenario, a given vehicle may be expected to arrive and clear the intersection prior to the next signal phase, assuming that the vehicle maintains its current speed. In this scenario, the simple solution is to ensure that the vehicle accelerates to the maximum allowable

speed in order to allow as many vehicles as possible to cross the intersection. Thus, this condition may be defined by the inequality:

$$t_{green\ remaining} > \frac{x_{vehicle} - x_{stopline}}{v_{vehicle}}$$

Should the inequality be false, the vehicle is expected to arrive after the signal turns red. As such, assuming that the vehicle is unable to speed up to clear the intersection, it would be best for it to slow down in order to reduce the shockwave area. A signal may be sent to the connected vehicle in question, informed it ahead of time that it cannot cross the intersection in time, allowing for more prudent driving behavior.

3.1.3.2 Scenario B: Vehicle Arrives at Intersection During Red Phase

In the second scenario, a vehicle may enter the sensor range during the red phase. If the vehicle is expected to arrive during the red phase, then it would be best for it to slow down ahead of time given the fact that the vehicle would have to wait at the intersection. This behavior would similarly reduce the shockwave area and result in a safer driving experience. The speed at which it slows down must be within a given threshold, as defined in the paper by Li et al. 2014 (Z. Li et al., 2014), as to not create additional shockwaves upstream. This scenario is thus given by the following inequality after which a “slow down speed” must be provided to the given vehicle:

$$t_{red\ remaining} > \frac{x_{vehicle} - x_{stopline}}{v_{vehicle}}$$

The vehicle may maintain its current speed rather than slow down should the current phase for the approach be red and the vehicle is expected to arrive once the signal changes to green, as given by:

$$t_{red\ remaining} < \frac{x_{vehicle} - x_{stopline}}{v_{vehicle}}$$

3.1.4 Defining the Proposed System

Considering the idealized system, a mixed dynamic programming and reinforcement learning approach was selected. In order to facilitate the task and reduce the exploration space, a set of conditions were established in which the system would apply ideal speeds obtained using a reinforcement learning algorithm. A flowchart of the proposed system is shown in Figure 2.

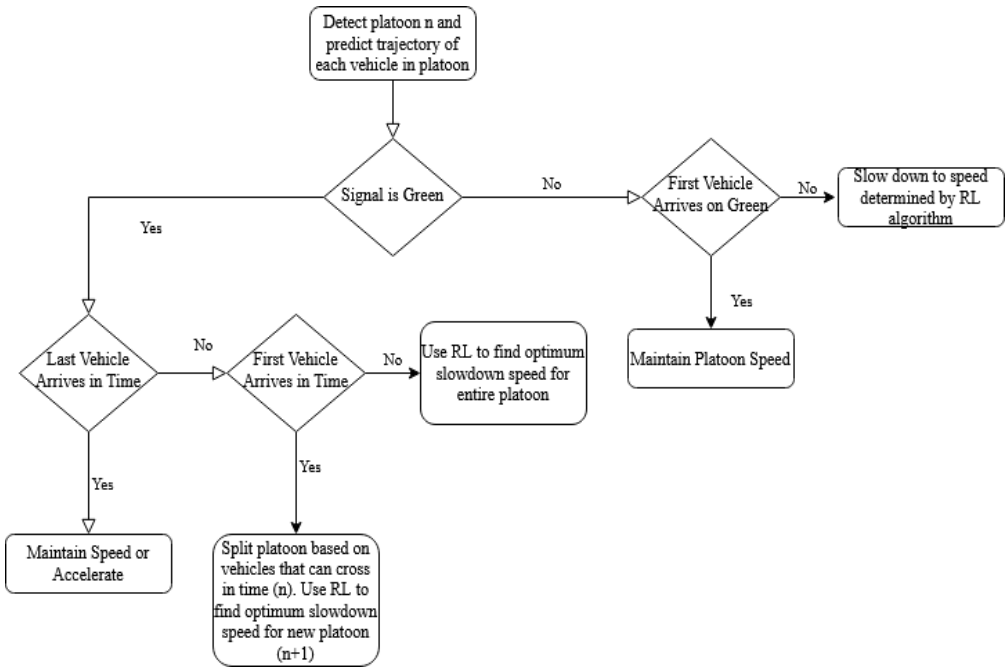

Figure 2: Flowchart Depicting a Simplified Version of the Algorithm.

Considering any given approach, the reinforcement learning algorithm would only activate if the system identified that a vehicle is unable to clear the intersection. Under these conditions, ideal slow-down speeds will be determined and applied to specific platoons in the approach. Simple solutions to the optimum behavior were identified, namely a situation where a vehicle may clear the intersection while traveling at maximum speed, and a situation where stopped vehicles prepare to clear the intersection as fast as possible. In the first situation, the simple solution would be to not slow down and clear the intersection to allow for fewer vehicles to be stopped once the signal phase changes. Similarly, in a scenario where vehicles are not in motion or are moving at

extremely low speeds (<5 km/hr) at a red light, the ideal behavior would be to accelerate to the speed limit once the phase changes, if it is safe to do so.

It should be noted that the algorithm deals with connected vehicles rather than autonomous vehicles. The speeds in question act as speed limits rather than actuated speeds. Consequently, issuing a speed advisory of 50 km/hr to a stopped vehicle informs a vehicle that it may travel up to 50 km/hr rather than informing it to accelerate to that speed immediately and potentially hit a stopped vehicle in front of it. Thus, the working assumption is that within the speed distribution commonly observed when setting a speed limit, road users tend to accelerate to the maximum legal driving speed on average and prefer higher speeds.

3.1.5 Determining the Ideal Speed

The ideal speed of a platoon of vehicles at any given point in time, v_{cr}, is dependent on the current state. Depending on the situations outlined in previous sections, the vehicle may be asked to speed up or slow down. While it may be possible to develop a formula using optimization which would eventually integrate into an ATSC system, this would be too computationally intensive given the large number of computations performed for every vehicle in the system. Instead, a reinforcement learning approach was used, provided that the situation warrants a slowdown. Reinforcement learning (RL) approaches are functionally similar to linear regression models with regard to their relatively low computational intensity. RL approaches are thus superior to conventional optimization approaches as they can also easily perform actions in real-time without the need to draft site-specific sets of equations, instead opting for a flexible and data-driven approach to achieve great safety benefits.

3.1.5.1 Network Architecture

Given the constraints relating to real-time use, reinforcement learning was considered to be the best approach. Due to the continuous state and action space, a simple Q-learning or Deep-

Q-Network approach was not feasible. As such, a model with a state space that may be discretized, and a continuous action space was selected. A DDPG architecture was thus used in this scenario and implemented using the MATLAB Reinforcement Learning Toolbox with three hidden layers of 100 neurons representing each of the actor and critic networks. 46 input nodes and 8 output nodes, representing the input variables and the actions respectively, were used.

DDPG agents use an actor and critic network in order to learn from estimated targets and upgrade the network to ensure stability. Overall, four neural networks were used including a Q-Network, a deterministic policy network, a target Q-Network, and a target policy network. Learning is characterized by maximizing the Bellman Equation, a formula used in reinforcement learning to describe the ideal action-state pair which optimizes for a given Q-function. Since a DDPG approximates this Q-function, it utilizes a mean-squared Bellman error function which is to be minimized given by: (Lillicrap et al., 2015; Silver et al., 2014)

$$L(\Phi, D) = E_{(s,a,r,s',d)\sim D}\left[\left(Q_{\Phi}(s,a) - \left(r + \gamma(1-d)Q_{\Phi targ}\left(s', \mu_{\theta_{targ}}(s')\right)\right)\right)^2\right]$$

A deterministic policy is selected which maximizes Q_Φ (s,a) via gradient ascent, as given by:

$$\max_{\theta} \underset{s\sim D}{\mathrm{E}}\, Q_{\Phi}\left(s, \mu_{\theta_{targ}}(s)\right)$$

Over time, the algorithm refines the Q-Network by updating the various weights resulting in improved performance. While the model is deterministic in nature, it utilizes an Ornstein Uhlenbeck noise function to account for random exploration while it exploits existing knowledge. The noise parameter used was assumed to be 10% of the total action space range with the predefined standard mean attraction rate of 0.15. Noise over time similarly decreases at a predefined rate provided that the Ornstein Uhlenbeck function is stable:

$$1 \geq |1 - \mu_{attraction} * t_{sample}|$$

The reward function, arguably the most important part of a reinforcement learning algorithm, was selected to be the sum of shockwave areas, which is the main parameter used to approximate traffic conflicts. The shockwave area was used in lieu of the other formulas which utilize platoon ratio. This was done since the safety performance functions obtained apply to entire cycles, and incremental changes in area cannot be represented as changes in conflicts given the significantly smaller timesteps. In theory, optimizing to reduce the shockwave area with several steps of look-ahead would be similar to optimizing for the exponential function of Shockwave Area obtained from Essa et al. 2018 (Essa & Sayed, 2018). This assumption was subsequently verified by the performance metrics used in the evaluation.

$$E(Y) = -\sum_{App=1}^{4} \sum_{Lane=1}^{Lane_n} V^{0.706} exp(-1.797 + 0.501\, A)$$

3.1.5.2 Defining the Action Space

The ideal speed, defined as the required speed to optimize the reward function in previous sections, is to be selected using a reinforcement learning DDPG approach. Every cycle, an unknown number of vehicles enter the intersection's detection range with anywhere from approximately 1 to 300 vehicles being detected by the system at a time. Selecting an individualized speed for each vehicle, while more optimal, is extremely difficult computationally considering that each output value represents a discrete speed.

To ensure that this system is applicable to real-world conditions, a sensor range of 225 m from the stop line was assumed to be the maximum range at which vehicles may transmit and receive information. This figure was obtained from the average ranges of standard V2I DSRC systems, which vary from 150 m to 300 m (U.S Department of Transportation, 2015). As a result of this range, there would likely be at most three platoons in the intersection at any given point in time,

with two platoons in motion and one stopped at the intersection. At all other points in time, a maximum of two moving platoons would likely be present given the spatial constraints.

Thus, for each approach, the leading vehicle speeds would be defined as $V_{platoon\,1}$ and $V_{platoon\,2}$. The desirable speed would be applied for all connected vehicles in the platoon. The range of allowable slowdown speeds selected was bounded between 30 km/hr and 50 km/hr, further simplifying the action space. For an intersection with four approaches, this represents 8 possible actions that may be taken. Each discrete speed within this range was assigned a speed distribution with a similar variance to that of a typical speed limit, but with a shifted mean. For example, a command requesting that vehicles slow down to 32 km/hr would yield a speed distribution resembling that of a speed limit of 30 km/hr with the mean shifted by 2 km/hr.

3.1.5.3 Defining the Input Variables

The state of an intersection from cycle to cycle is continuous and a near-infinite number of variables with countless combinations may be used. By examining the literature and observing key parameters, it was determined that this continuous state space should be discretized. Variables were obtained for each of the 4 approaches ensuring that said variables were readily available via V2V and V2I data. Thus, a set of 4 inputs was obtained for each variable, each representing the state of their respective approaches.

The number of vehicles occupying a specific portion of the road was among the most important variables used. The number of vehicles would theoretically be key in providing the system with an understanding of the state, as typically fewer vehicles on the road result in fewer traffic conflicts. However, with trajectory optimization, the distribution of vehicles would be equally as important. As such, the roadway may be split into four different segments, each with a certain number of vehicles as shown in Figure 3. Thus, segments of 56.25 m were selected to represent occupancy and the spatial distribution of vehicles. While it would have been possible to

further split the roadway into smaller segments, this would result in additional variables which may complicate the convergence process. In the example shown in Figure 3, the system would register one vehicle in Section 1, five in Section 2, four in Section 3, and three in Section 4, registering their respective speeds. In this Figure, blue vehicles represent connected vehicles, and the larger black vehicle represents a bus. By considering the spatial distribution of roadway occupancy, these variables help prevent the system from selecting actions that would adversely affect platoon length and, by extension, safety.

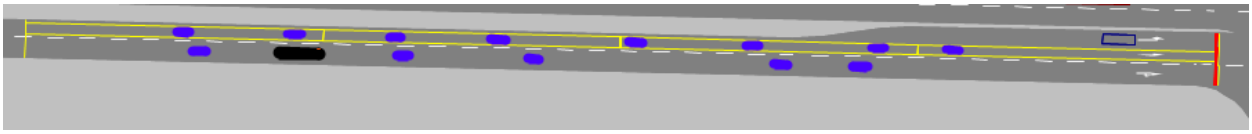

Figure 3: Graphical Representation of the Roadway in VISSIM Split into Four Parts on One Lane (n1=1, n2=5, n3=4, n4 = 3).

Average speed was also selected as a significant variable. While platoon speed could have been used, average speed allows the system to recognize where vehicles are slowing down and where they are accelerating. Average platoon speed would be restrictive with respect to allowing the system to combine or divide different platoons and would not be conducive towards identifying stop-and-go waves. A similar methodology was used to account for the spatial distribution of vehicle speeds. The same 4 roadway segments were utilized, and an average speed was obtained for each segment.

In order to complement the spatial variables selected and to capitalize on the benefits of platooning, the gap between subsequent platoons was selected as an input variable. Given the sensor range of 225 m, there would typically be at most three platoons within range at any given point in time. For the algorithm, merging the platoons would require some degree of awareness of the distance between subsequent groups of vehicles. Thus, the gap between the first and the second

platoon of vehicles was selected as a key variable. If a second or third platoon does not exist, the distance between the last detected platoon and the sensor range is assumed to be the platoon gap to overcome the limitations of the sensor range. The system assumes that if there is no detected platoon in range, a platoon will arrive shortly thereafter.

Temporal variables are equally important in trajectory optimization. Neural networks typically do not work well with binary variables since fractions or strange weights may arise during the backpropagation process and in the computation of errors. To overcome this limitation, rather than represent the signal states as a binary value, the current time elapsed since a given phase is a superior measure to inform the system about how much time is left in the current signal phase on a given approach. Based on this information, higher or lower speeds may be selected depending on whether there will soon be a phase change. Consequentially, the “time since last red” and “time since last green” were used as temporal variables.

It should be noted that the system optimizes all four approaches simultaneously. As such, it may be difficult for the system to identify which variable corresponds to which approach resulting in some non-existent correlations being erroneously detected. A collaborative multi-agent approach may be used to allow for four DDPG agents to simultaneously optimize each approach, however, this may complicate matters when adding additional dimensions to the problem such as signal optimization. A simpler solution is to allow the output of the precedent action to be represented as an input to the following computation. This allows for a single agent system to be able to better identify the rewards resulting from a given action on a given approach. The explanatory variables selected as inputs to the model are detailed in Table 1.

Table 1: Relevant Input Variables Considered by the Reinforcement Learning Algorithm.

Variable Name	Definition	Variable Name	Definition
n11	Number of Vehicles on Approach 1 Occupying Region 1	v32	Average Speed of Vehicles on Approach 3 Occupying Region 2
n12	Number of Vehicles on Approach 1 Occupying Region 2	v33	Average Speed of Vehicles on Approach 3 Occupying Region 3
n13	Number of Vehicles on Approach 1 Occupying Region 3	v34	Average Speed of Vehicles on Approach 3 Occupying Region 4
n14	Number of Vehicles on Approach 1 Occupying Region 4	PGap3	Gap between first and second platoon on Approach 3
v11	Average Speed of Vehicles on Approach 1 Occupying Region 1	n41	Number of Vehicles on Approach 4 Occupying Region 1
v12	Average Speed of Vehicles on Approach 1 Occupying Region 2	n42	Number of Vehicles on Approach 4 Occupying Region 2
v13	Average Speed of Vehicles on Approach 1 Occupying Region 3	n43	Number of Vehicles on Approach 4 Occupying Region 3
v14	Average Speed of Vehicles on Approach 1 Occupying Region 4	n44	Number of Vehicles on Approach 4 Occupying Region 4
PGap1	Gap between first and second platoon on Approach 1	v41	Average Speed of Vehicles on Approach 4 Occupying Region 1
n21	Number of Vehicles on Approach 2 Occupying Region 1	v42	Average Speed of Vehicles on Approach 4 Occupying Region 2
n22	Number of Vehicles on Approach 2 Occupying Region 2	v43	Average Speed of Vehicles on Approach 4 Occupying Region 3
n23	Number of Vehicles on Approach 2 Occupying Region 3	v44	Average Speed of Vehicles on Approach 4 Occupying Region 4

Variable Name	Definition	Variable Name	Definition
n24	Number of Vehicles on Approach 2 Occupying Region 4	PGap4	Gap between first and second platoon on Approach 4
v21	Average Speed of Vehicles on Approach 2 Occupying Region 1	TSG	Elapsed Time Since Green Phase on Approach 1
v22	Average Speed of Vehicles on Approach 2 Occupying Region 2	TSR	Elapsed Time Since Red Phase on Approach 1
v23	Average Speed of Vehicles on Approach 2 Occupying Region 3	A11	Previous Action representing the input speed for Platoon 1 on Approach 1
v24	Average Speed of Vehicles on Approach 2 Occupying Region 4	A12	Previous Action representing the input speed for Platoon 2 on Approach 2
PGap2	Gap between first and second platoon on Approach 3	A21	Previous Action representing the input speed for Platoon 1 on Approach 2
n31	Number of Vehicles on Approach 3 Occupying Region 1	A22	Previous Action representing the input speed for Platoon 2 on Approach 2
n32	Number of Vehicles on Approach 3 Occupying Region 2	A31	Previous Action representing the input speed for Platoon 1 on Approach 3
n33	Number of Vehicles on Approach 3 Occupying Region 3	A32	Previous Action representing the input speed for Platoon 2 on Approach 3
n34	Number of Vehicles on Approach 3 Occupying Region 4	A41	Previous Action representing the input speed for Platoon 1 on Approach 4
v31	Average Speed of Vehicles on Approach 3 Occupying Region 1	A42	Previous Action representing the input speed for Platoon 2 on Approach 4

3.2 Validation and Testing

3.2.1 Modeling the Environment

The proposed model was applied to a simulation to determine the potential benefits of the system. VISSIM simulation software was used to model the effects of this framework on a real intersection. Data from two intersections, 128th Street & 72nd Avenue, and 132nd Street & 72nd

Avenue in Surrey BC, were utilized to create and calibrate a VISSIM model and ensure an accurate representation of reality (Essa & Sayed, 2018).

The modeling environment, PTV VISSIM, utilizes a Wiederman 99 car-following model. The model, which accurately represents car following behavior, allows for the creation of several different vehicle classes. A custom class relating to connected vehicles was created which allows for all V2I and I2V communications to be limited to vehicles of the CV class.

PTV VISSIM's COM integration allows for the extraction of real-time traffic data from the simulation including position, speed and other relevant parameters, and dynamic assignment of desired speed distributions. Different parameters may also be changed during a simulation including signal states as well as individual kinematic variables for specific vehicles. Of note is the desired speed of a vehicle, which may be modified during the simulation. The desired speed represents a predefined speed distribution that defines a vehicle's acceleration and deceleration behavior. This may be used to emulate speed advisories, given that the behavior is similar to driver reaction to speed limits, and slight variations in speeding/slowing down behavior are likely to be observed in practice. Speed distributions were created in PTV VISSIM emulating the typical distribution of vehicle speeds at the speed limits of 30, 40, and 50 km/hr with a shifted mean to represent desired speed distributions for discrete speed limits within this range (e.g. centred at 32 km/hr, 36 km/hr, 42 km/hr, 49 km/hr, etc.).

3.2.2 Defining the Test Sites

A total of two intersections were modeled with varying scenarios created to represent different arrival patterns and traffic volumes throughout the day. The first intersection on 128th Street and 72nd Avenue was used to train the reinforcement learning model while the second intersection was set aside to be used to validate the results.

The first intersection, shown in Figure 4, is located on 128th Street and 72nd Avenue in Surrey, BC, is comprised of four approaches, each with two through lanes and a left turn lane at the intersection. A bus route is present on 72nd Avenue. A loop detector system that allows for adaptive traffic signal control using left turn vehicle occupancy was also present on all four approaches. This was considered as the main test site which represents a typical intersection. The algorithm's performance on this intersection was considered to be representative of standard conditions for two-lane roads.

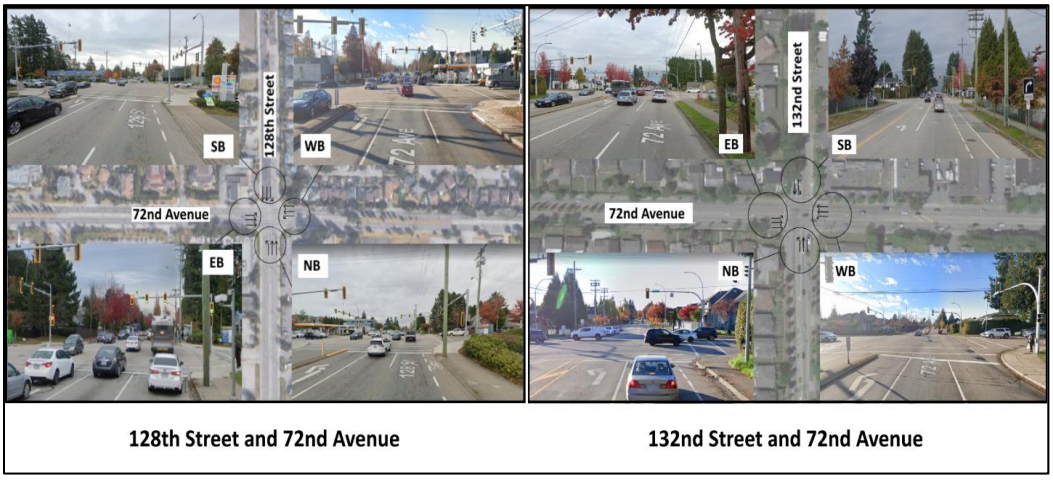

Figure 4: 128th Street and 72nd Avenue and 132nd Street and 72nd Avenue.

A second intersection located on 132nd Street and 72nd Avenue in Surrey, BC was selected to validate the algorithm due to its distinct characteristics. Unlike the first intersection, this intersection features a southbound approach with one lane and a left-turning bay, as well as a northbound approach with a right turning lane, a through lane, and a left-turning bay. The eastbound and westbound approaches are similar to that of the first intersection, with two through lanes and a left turn lane on each approach. Detector loops and bus routes were present on all four approaches of this intersection. These differing characteristics allow for the algorithm to be tested on various different geometric and temporal conditions to provide insight to the effectiveness of the algorithm on different types of roadways. An additional benefit of using a secondary

intersection as validation would be to observe the effects of transferring a trained model onto a new intersection and observing the effect of training.

The hourly traffic volumes observed on 128th Street and 72nd Avenue ranged from 2,343 vehicles at 10:00 AM - 11:00 AM to 4,186 vehicles at 3:00 PM - 4:00 PM, and the hourly traffic volume observed on 132nd Street and 72nd Avenue ranged from 2,140 vehicles at 10:00 AM - 11:00 AM to 3,354 vehicles at 5:00 PM - 6:00 PM. It should be noted that the southbound and northbound traffic volumes are for a single lane of traffic which differs from the two-lane northbound and southbound approaches on 128th Street. The hourly traffic volume data are shown in Figures 5 and 6.

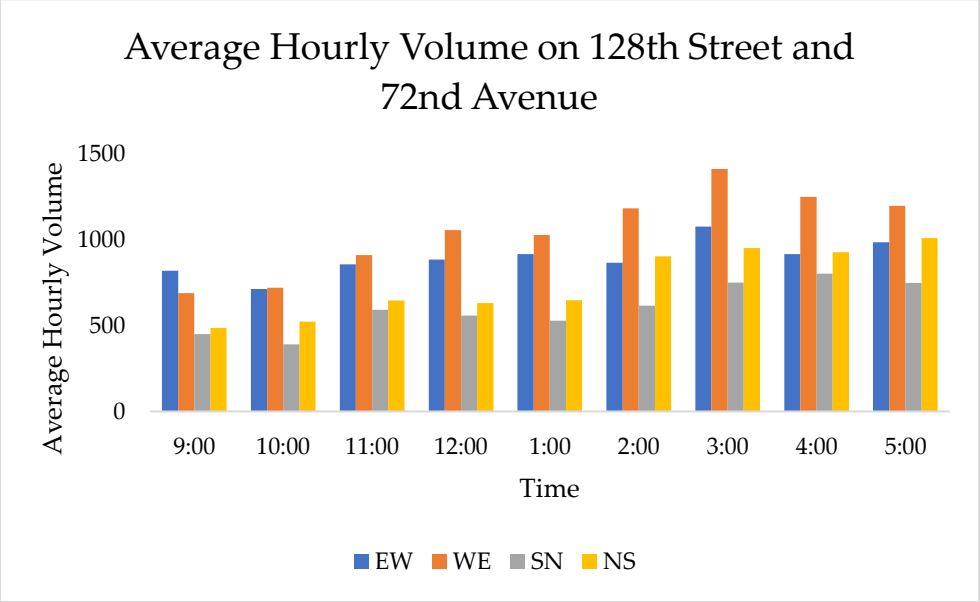

Figure 5: Hourly Average Traffic Volumes on 128th Street and 72nd Avenue.

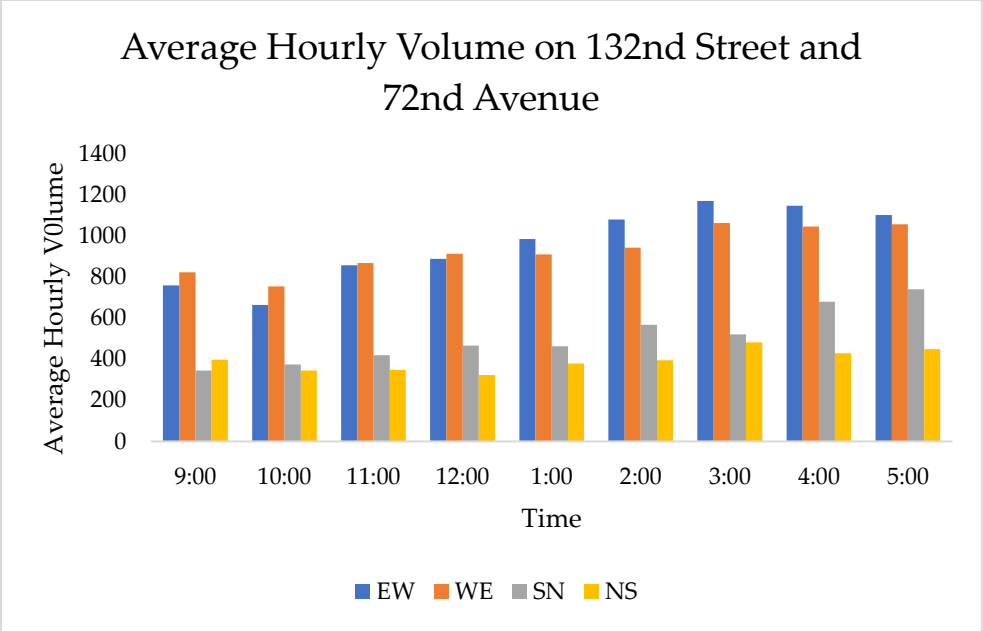

Figure 6: Hourly Average Traffic Volumes on 132nd Street and 72nd Avenue.

3.2.3 Training Process

The first intersection at 128th Street and 72nd Avenue in Surrey, BC was used to train the model. Episodes, which consisted of random 20-minute samples from simulation hours, were used in the training process during which the model's performance was expected to improve over time. The model was initially trained on traffic conditions which represent the scenario between 12:00 PM and 1:00 PM, before being applied to random hours between 9:00 AM and 5:00 PM. This was to generalize the training dataset as to not overfit to a specific traffic volume. Over 450 episodes were run, with a minimum of 50 episodes per simulation hour being performed for each scenario. While simulation time may vary based on the device used, each episode took approximately an average of 8 minutes to complete, for a total approximate training time of 60 hrs.

Each simulation episode consisted of several timesteps, with a defined Δt of 5 seconds. The trajectories of vehicles were obtained every second in real-time. These trajectories were used to update a trajectory matrix every Δt, which was used to obtain both the input parameters to the RL model (Occupancy, Speed, Platoon Gap, etc.) and to predict future trajectories. The system

thus considers these future trajectories when updating the platoon speed and relays speed advisories to the vehicles in question through the use of desired speed distributions.

VISSIM random number generator (RNG) seeds were used to account for the inherent randomness in the system to overcome limitations within the model and to avoid overfitting using the DDPG RL algorithm. The Ornstein Uhlenbeck noise parameters were similarly selected to avoid repeatedly applying the same "optimum policy" to the same situation without exploring potentially better solutions and updating the policy based on new states observed.

Once the system was trained on the intersection at 128th Street and 72nd Avenue for 450 episodes, baseline safety performance was established based on fixed seeds throughout the day. This established the basis for which model performance could be compared. The process was repeated for the second intersection resulting in the creation of a "do-nothing option" scenario where baseline safety was quantified.

3.2.4 Testing Procedure

After 450 episodes of training, the model was run for an additional 100 episodes to obtain an average number of conflicts per approach. These results were then compared to the baseline simulations where no actions were being taken.

To ensure transferability, a secondary intersection at 132nd Street and 72nd Avenue was used to validate the model and ensure that the benefits are indeed transferable without additional training and applicable to approaches with different characteristics.

Initially, the model was run assuming that all vehicles on the road are CVs that comply with speed advisories. It is not practical to assume 100% MPR in the near future, and hypothetically even if it is achieved, compliance remains an issue. However, this scenario represents the maximum possible improvement as a result of using this system. Alternate scenarios relating to a mixed environment were created and tested. Scenarios with 100%, 75%, 50%, 25%,

15%, and 10% market penetration ratio were tested to examine the effects of MPR, and by extension low compliance. Conventional vehicles are assumed to be invisible for the purpose of the RL algorithm inputs and do not receive speed advisories. They are nonetheless considered in the final computation of traffic conflicts. MPRs under 10% were not tested due to a high expected variance, which would be more reflective of the simulation environment than real-world conditions. Thus, 10% MPR was considered to be reliable enough to be used to estimate system performance at lower levels of MPR.

The performance of over 100 trial runs were observed for each scenario, and the distribution of results was compared based on each run's conflict reduction to evaluate performance. Given that each intersection consists of 4 different approaches and the system was tested on two intersections, the system was tested on a total of 8 different independent approaches, each with 6 different fleet compositions. Since the system tests an individual controller for all four approaches of an intersection, the overall net safety benefit was considered in the computation of the results rather than individual benefits to each approach. Randomness during the training process was accounted for through the use of the initial RNG seed built into the VISSIM simulation as well as a series of dummy intersections simulating an arrival pattern representative of the study intersections.

The training process was based on expected real-world behavior. Given that the system uses a DDPG algorithm to train the model, there is a large amount of inherent randomness in the first few sets of exploratory actions. As such, a completely untrained model may perform poorly in practice and result in additional traffic conflicts that could have been prevented. To remedy this, the intent of this training process was to create a model that may be trained on a real or simulated intersection that could then be transferred onto another similar intersection to achieve some safety

benefits at the new intersection without additional training. Additional training would further improve safety at that intersection over time as the system learns more intersection-specific patterns and characteristics.

3.3. Results

3.3.1 Quantifying Safety at Full Implementation

After running hundreds of simulations, the algorithm managed to yield an average conflict reduction of 23% when applied to 128th Street and 72nd Avenue. The overall results were promising, with most trials seeing a statistically significant improvement to safety ranging from a minimum daily conflict reduction of -2% to a maximum of 38% when considering the sum of best and worst-case scenarios for every hour. The statistical parameters shown in Table 2 represent the sum of hourly conflicts compared to the total sum of hourly baseline conflicts between 9:00 AM and 5:00 PM:

Table 2: Projected Safety Benefits on 128th Street and 72nd Avenue.

Parameter	Value
Average Hourly Conflict Reduction	23%
Median Hourly Conflict Reduction	20%
Sum of Maximum Hourly Conflict Reduction (Best Case Scenario)	38%
Standard Deviation	10%

The implementation of this algorithm at full market penetration without considering additional benefits as a result of other technologies available with CVs and CAVs yields considerable results. On an hourly basis, an improvement to safety may be expected 88% of the time. When considering the sum of worst-case scenarios for every hour, a daily increase in traffic

conflicts of 2% may be expected, indicating that any negative effects observed in one given hour will likely be offset by benefits in another.

The algorithm was tested on the test intersection over the course of a day with traffic conditions between 9:00 AM and 5:00 PM being modeled in VISSIM. The results ranged from significant reductions in hourly traffic conflicts to negligible reductions based on the time of day. Initially, a compensation effect was observed where approaches with less exposure would achieve comparably less improvement in absolute due to there being fewer conflicts to prevent. However, as the traffic volume increased, greater improvements were observed. This was highly dependent on the arrival patterns in which certain more favorable patterns resulted in greater conflict reduction as seen in Figure 7. Traffic volume did not seem to have a considerable effect on percent conflict reduction.

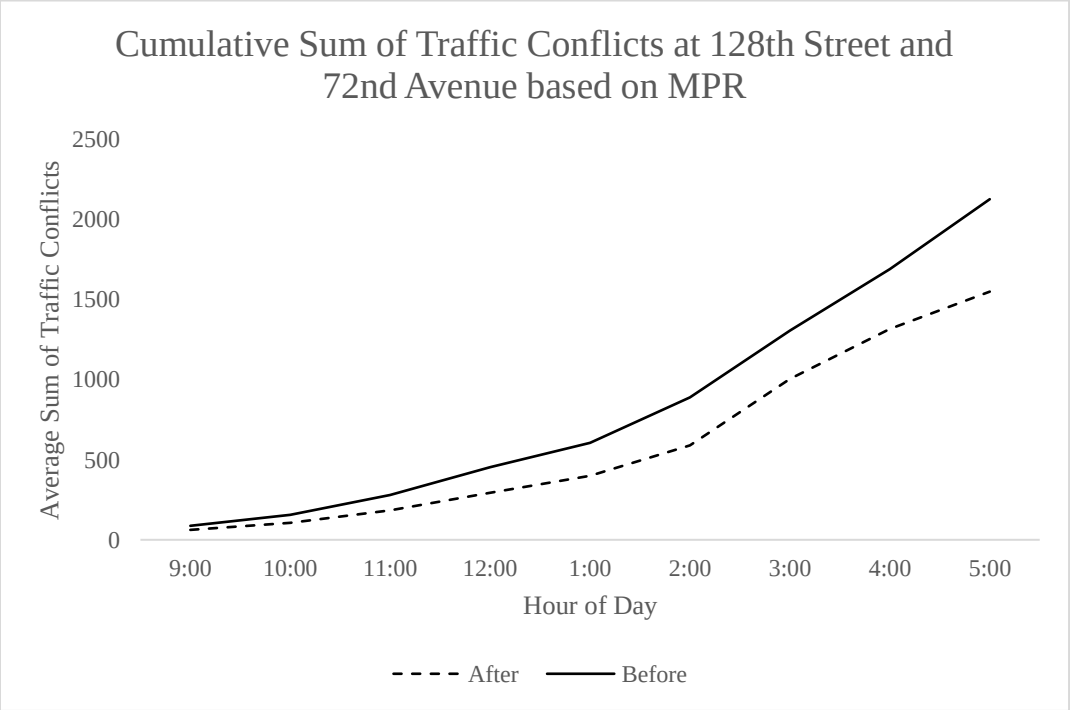

Figure 7: Variation of Conflict Rate per Hour on 128th Street and 72nd Avenue.

3.3.2 Impact of MPR on Safety

Test scenarios for 100%, 75%, 50%, 25%, 15%, 10%, and 0% MPR scenarios were tested on 128th Street and 72nd Avenue and yielded positive results. Similar to previous studies, the bulk of the safety benefits begin to manifest when approximately a quarter of vehicles on the road are CVs, with the largest jump in conflict reduction observed between 15% and 25% MPR where 74% of the total benefits are attained. At 10% MPR, a 6% total reduction in conflicts can be observed, indicating that the system can improve safety early in the CV implementation process.

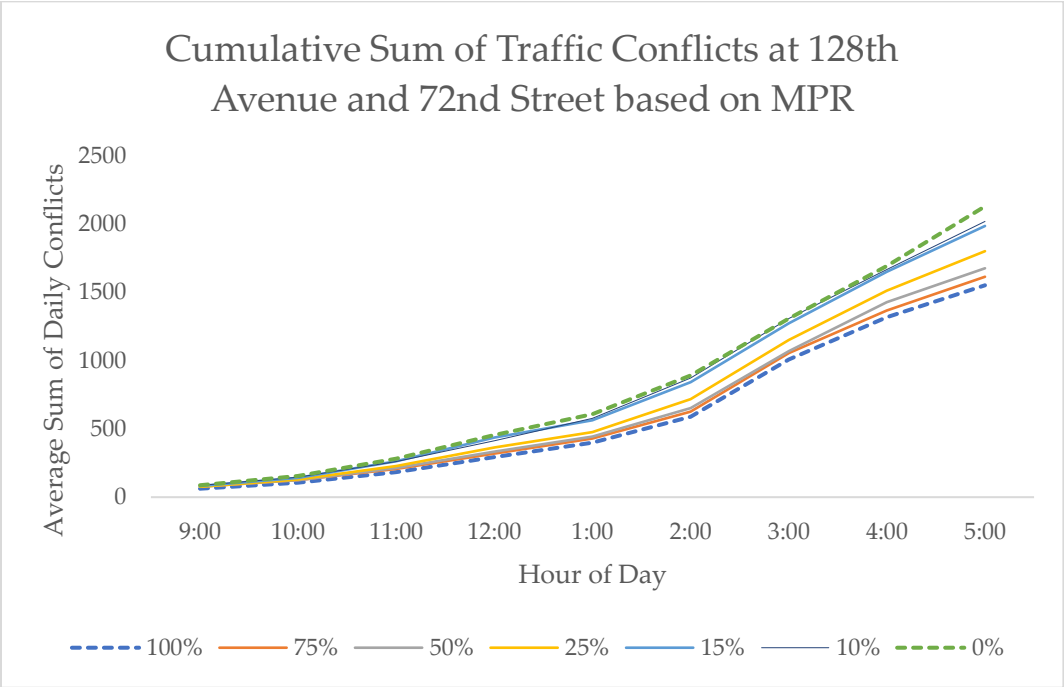

Figure 8: The Effect of Market Penetration Rate on Conflict Rate.

This indicates that this system is functional in the transitional state when not all vehicles on the road are connected and thus is applicable in the near future. However, certain outliers were observed, as seen in Figure 8, which resulted in a higher MPR corresponding to increased conflicts at lower levels during certain times of the day. This indicates that the system may work in most scenarios but underperforms under certain conditions. A compensating effect is however observed when considering the entire dataset which offsets the negative effects.

3.3.3 Validation and Transferability

To ensure that the system is transferable, the same algorithm was applied to 132nd Street and 72nd Avenue without additional training. 132nd Street and 72nd Avenue have different roadway characteristics including two one-lane through approaches (as opposed to two-lane approaches) and a far larger left turn bay and left-turn volume.

Without training and including all approaches, the total daily improvement to safety was 9%, varying greatly throughout the day. The system reduced an average of 61 conflicts per hour, with a peak conflict reduction of 412 conflicts at 3:00 PM and mixed performance throughout the day. When comparing the intersections and controlling for volume, the peak conflict reduction at 132nd Street & 72nd Avenue was 28% at 3:00 PM as compared to a peak reduction of 42% at 5:00 PM on 128th Street & 72nd Avenue. The average daily reduction in traffic conflicts on 128th Street & 72nd Avenue and 132nd Street & 72nd Avenue were 23% and 9% respectively. This indicates that the algorithm yielded positive values on the validation intersection, but not to the same degree as the benefits seen on the test intersection.

The disparity in the observed benefits may have been attributed to the difference in geometric characteristics of the various approaches. To ensure that an appropriate comparison was made, the results were filtered to only account for similar approaches (those on 72nd Avenue). As a result, a total average conflict reduction of 15% was observed as shown in Figure 9 and Table 3.

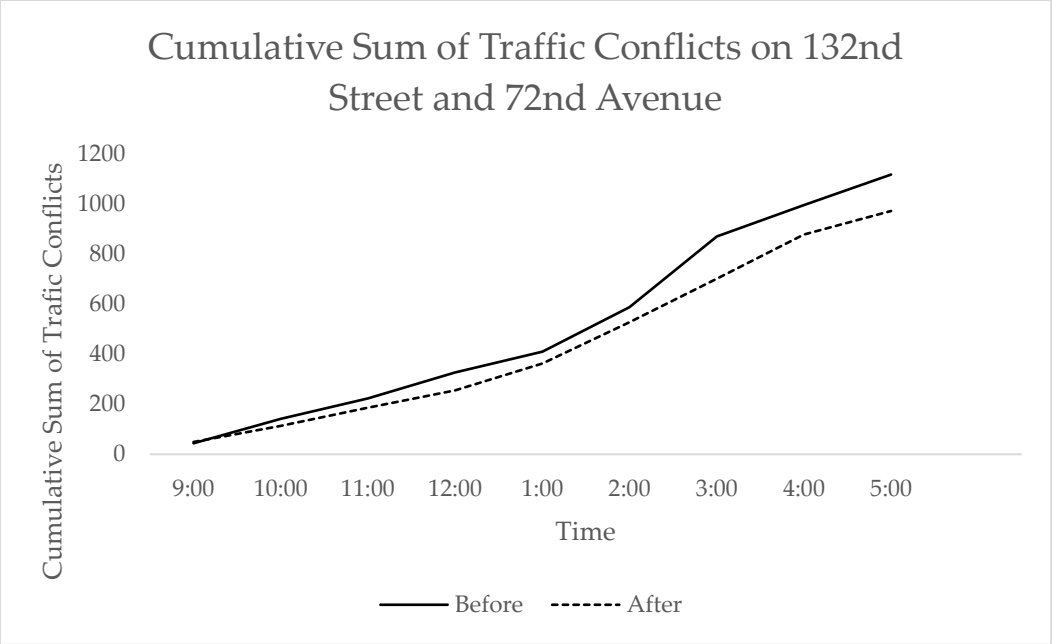

Figure 9: The Effect of Market Penetration Rate on Conflict Rate.

Table 3: Variation of Conflict Rate per Hour on 72nd Avenue (EB and WB Approaches).

	Entire Intersection	**EB and WB Approaches**
Average Hourly Conflict Reduction	9%	15%
Standard Deviation	9%	9%

Despite the different geometric and temporal characteristics of 132nd Street, it can be said that the proposed system would result in a net improvement in road safety without training. With additional training, the benefit would be significantly greater and similar effects may be observed. However, this serves as an important visualization of what would happen if the trained dataset were not fully representative of the spatial conditions at another intersection. This is not to say that this algorithm is ineffective in other intersections, rather that site-specific algorithm training is required to maximize the benefits.

3.4 Discussion

3.4.1 Observed Vehicle Behaviour

The proposed algorithm demonstrated behavior similar to the ideal behavior previously discussed, with vehicles slowing down such that they may arrive at the intersection in time for the next green while avoiding the creation of stop-and-go waves or unnecessary delays. The space-time diagram in Figure 10 demonstrates this relationship where a smaller shockwave area was observed over eight cycles due to the algorithm's ability to dissipate shockwaves by controlling individual vehicle speeds. It should be noted that this represents the trajectory observed on a single lane, and that lane changing behavior may influence the visualization creating the discontinuities observed in Figure 10.

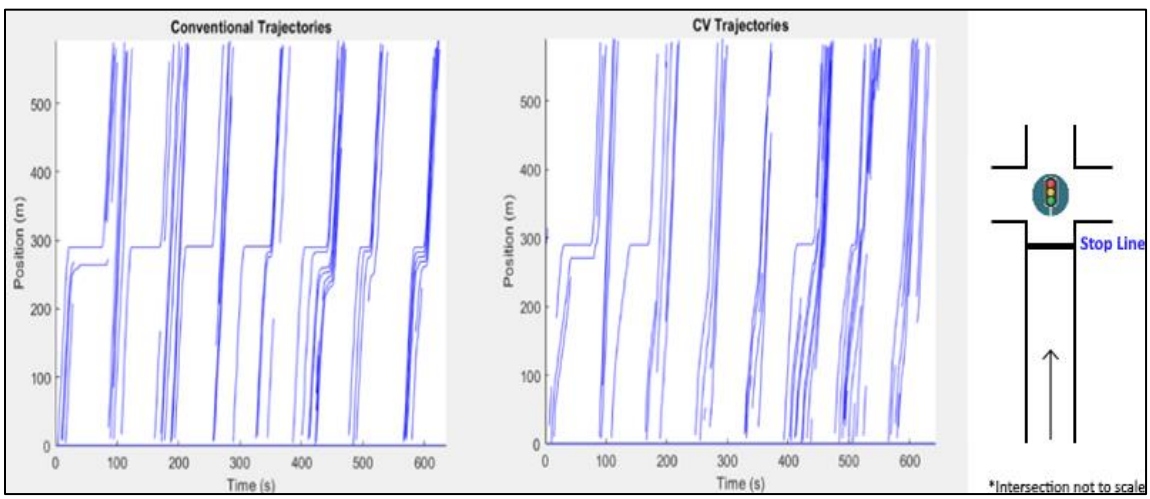

Figure 10: Side by Side Comparison of Space-Time Diagrams for the Westbound Approach on 72nd Avenue (128th Street and 72nd Avenue) (Stop Line at x=300 m).

This behavior also applies to scenarios without full market penetration. In which case, the effectiveness of the slowdown commands was dictated primarily by the probability of a connected vehicle being the leading vehicle of a platoon. In situations where the leading vehicle or subsequent vehicles were not connected, the benefit was not seen, and the shockwave was not as effectively dissipated. This indicates that while the full implementation condition results in consistent safety

benefits, the intermediate state would see significantly greater fluctuations in conflict reduction. At 100% MPR, the system appears to control the queue length without the need to set static upper bounds. In most cases, this prevents the formation of additional shockwaves upstream.

It is worth further analyzing the performance of 132nd Street, which did not perform as well as 128th Street. This is largely attributed to the differences based on the different types of approaches. On the eastbound and westbound approaches of 132nd Street and 72nd Avenue, the algorithm performed comparably to the approaches on 128th Street and 72nd Avenue due to the similar characteristics. A similar variance in the safety benefits was also observed at these intersections.

Further analyses were thus performed on the northbound and southbound approaches which had different constraints. Upon further observation, the bulk of the increase in traffic conflicts resulting from the implementation of the algorithm stemmed from the southbound approach. Since there is only one lane, right-turning vehicles, which may turn right on red, may reduce traffic flow if they reduce their speeds at a red light. Given the shorter signal time, these vehicles and slower moving traffic, in general, cause an accumulation of vehicles. Less vehicles are able to clear the intersection per cycle, resulting in increased conflicts due to longer queues. This offsets benefits gained on other approaches. Similarly, left-turning vehicles seeking to change lanes into the left turning bay may slow down right-turning vehicles. This phenomenon is not observed on the other approaches due to the ability of vehicles to change lanes and the significantly smaller right-turning volume from the minor road.

In contrast, this result was not seen on the northbound approach as right-turning vehicles merged into the right-turning lane and thus did not unnecessarily slow down following vehicles. The individual impact of this approach was estimated to have resulted in a 22% reduction in

conflicts relative to the baseline when compared to the southbound approach which resulted in a 34% increase in conflicts, offsetting the greater gains to safety elsewhere. These effects were further influenced by the difference in traffic volumes on each approach.

The implementation of this system is possible and recommended if prior studies are performed using data specific to an intersection. The algorithm functions well in both high volume and low volume conditions, provided that there are multiple lanes available and the right turning traffic is either exclusive, protected, or small relative to the through volume. Blind or inappropriate implementation of this algorithm, as with any other safety-related intervention, may worsen the situation if executed poorly. Nonetheless, the results show benefits in multi-lane intersections with lesser benefits on single-lane roads. The reinforcement learning aspect of the algorithm must be trained on the intersection in question. Should additional training be performed, the results are expected to drastically improve. As such different parameters related to roadway geometry, arrival patterns, vehicle compositions, and volume that are site-specific should be considered in the training model.

3.4.2 Market Penetration Rate and Implications for Policy

The proposed algorithm provides a low-cost, low computational intensity intervention that optimizes safety and reduces rear-end traffic conflicts. The conflict reduction scales with increased MPR and has diminishing returns at higher rates. 74% of the total benefit can be obtained when one in four vehicles has I2V and V2I capabilities, not including additional safety benefits resulting from the proliferation of connected vehicles.

While an average conflict reduction of 15-23% may be expected as a result of trajectory optimization alone, further improvements to safety may result from other factors such as the use of CVs in adaptive traffic signal control systems, and other systems which utilize CV data and CAV technologies to improve safety. Other technologies that are likely to proliferate alongside

CVs such as assisted braking and lane assist systems, would amplify the effects of this system and further improve safety. The most significant intervention to further improve safety are ATSC systems. ATSC systems may be used to dynamically optimize traffic signals to help dissipate shockwaves that form due to slowed or stopped traffic. When combined with trajectory optimization, significant benefits to safety are likely to be observed.

Chapter 4: Application of RL to SVCC Algorithm

This chapter explores a novel reinforcement learning-based SVCC system that utilizes an SAC algorithm to optimize for both vehicle speeds and signal timing. A variable signal time extension is established using safety performance functions and the traffic safety state is quantified such that long-term safety may be optimized. Vehicle trajectories are used to predict future trajectories, which are then modified using speed advisories in conjunction with signal time modification. The methodology and effectiveness of the system are explored in the rest of the section, and the benefits and limitations are discussed.

4.1 Developing a Signal-Vehicle Coupled Control System

This thesis seeks to develop a framework for a signal-vehicle coupled control system where speed advisories and signal extensions are jointly issued by the same controller. A hybrid dynamic programming and reinforcement learning approach was used to organize incoming traffic into platoons moving at an ideal speed and allocate additional signal time to these platoons. System performance was evaluated using a simulation model and compared to a baseline model, much like in the previous section. Real-world video data collected from intersections were used to model the test and validation intersections and quantify the expected benefits from using the proposed system.

4.1.1 Selecting a Function to Quantify Safety

The safety performance functions (Essa & Sayed, 2020) used to develop the algorithm discussed in Chapter 3 were used again to develop the proposed algorithm. Using dynamic traffic parameters such as shockwave characteristics, safety can be quantified as follows:

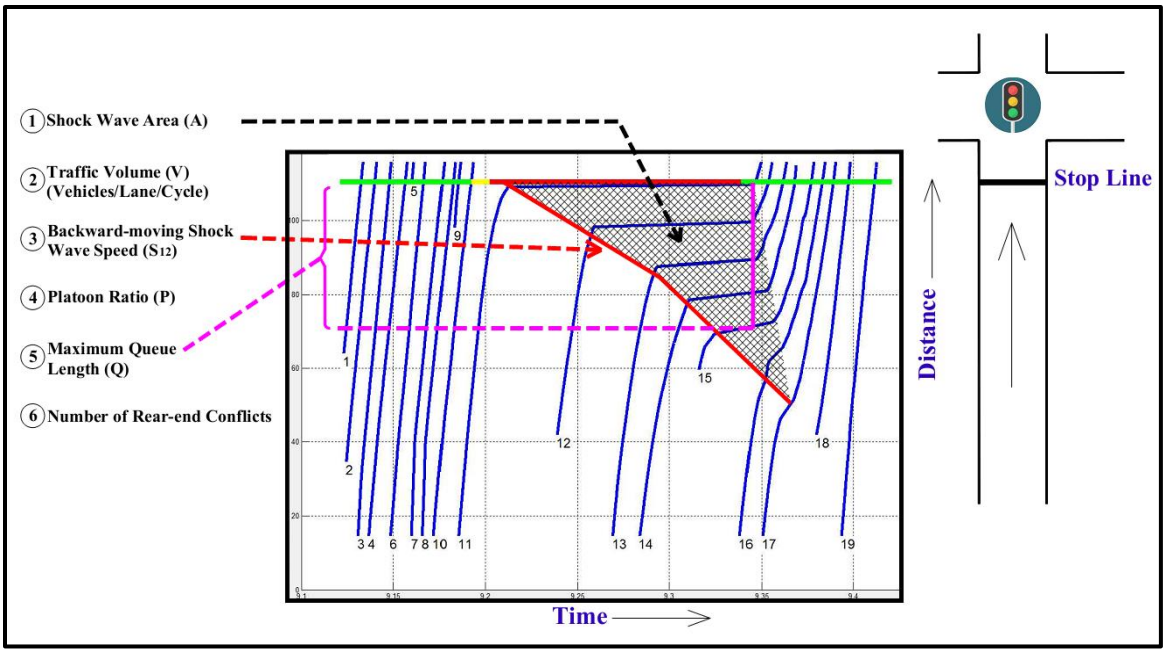

Figure 11: Shockwave Parameters and Traffic Conflict Models, (Essa & Sayed 2018)

1. $\boldsymbol{E(Y)} = V^{1.563} exp(-3.231)$

2. $\boldsymbol{E(Y)} = V^{0.706} exp(-1.797 + 0.501\ A)$

3. $\boldsymbol{E(Y)} = V^{0.65} exp(-2.046 + 0.0122\ Q)$

4. $\boldsymbol{E(Y)} = V^{1.637} exp(-3.316 + 0.05\ S_{12})$

5. $\boldsymbol{E(Y)} = V^{1.571} exp(-1.768 - 1.266\ P)$

6. $\boldsymbol{E(Y)} = V^{1.239} exp(-1.624 + 0.294\ A - 0.828\ P + 0.119\ S_{12})$

The models were derived using data from an entire traffic signal cycle on those intersections. Selecting timesteps that reflect the time intervals in which the data were collected would allow for a more accurate quantification of safety at the intersection. In the original study, Equation 6 had the highest explanatory power and as such was selected as the most appropriate function to quantify safety for the development of this algorithm. An entire phase is required to compute the variables for Equation 6. Consequently, a timestep was defined based on signal

phases. A conflict rate may thus be computed every phase, in contrast with the algorithm in Chapter 3 which considered timesteps of five seconds. This method is expected to yield a more accurate estimation of traffic conflicts. Safety as a performance measure for the purpose of this thesis can be defined as the sum of conflicts obtained from applying Equation 6 to every lane and approach.

It should be noted that these functions only consider rear-end conflicts. The reduction of rear-end conflicts using the proposed algorithm is unlikely to cause an increase in other types of conflicts. Nonetheless, the effect on alternative collision types such as left-turn conflicts and sideswipes are to be examined in future research.

4.1.2 Environment State Representation

To develop an algorithm that can optimize for safety, the state space should be well-defined and measurable. The objectives of the algorithm are to optimize signal timing and vehicle trajectories for any given signalized four-way intersection. Given an average direct short-range communication (DSRC) range of 225 m (U.S Department of Transportation, 2015), the environment space can be defined as four 225 m spans of roadway leading to an intersection. Both spatial and temporal characteristics are represented as the system is dynamic. The main components modeled are the presence of vehicles in the intersection, their kinematic parameters, and the signal timing.

A variety of temporal and spatial variables may be included in a detailed representation of the environmental state. A complete representation of the environment state would thus consider both the signal state and timing, as well as discrete vehicle driving behavior, vehicle type, kinematic parameters, and spatial occupancy at every given point in time. Given that there are hundreds of vehicles potentially within range of the DSRC system, it is not realistic to represent each vehicle individually. An aggregate representation that accurately approximates spatial and temporal variables should be used to address this issue. In the previous chapter, these variables

were each individually defined and represented in the model based on the location of a group of vehicles. In this section, an alternative approach is proposed using the safety performance functions. The added complexity of simultaneous signal timing optimization and vehicle trajectory optimization makes the approach used in Chatper 3 difficult to implement. Using the formulae derived in previous sections to approximate safety for each approach, the environment state can be adequately represented due to the various parameters which are considered in this computation.

First, the number of vehicles on an approach is considered a key variable that may be used to represent occupancy. While this does not necessarily indicate vehicle location, the cycle level traffic volume is correlated with conflict rate and when used in a dynamic model, can approximate the change in relative density and flow rates over time.

Platoon ratio, defined as the number of vehicles arriving on green on a given approach during a cycle, is another important measure. Platoon ratio serves as a surrogate measure for delay as the two variables are strongly correlated. A high platoon ratio represents conditions where the arrival patterns are favorable, and vehicles arrive in organized platoons. A reduction in the platoon ratio would be easily recognized by the system as unfavorable and thus avoided.

Shockwave parameters such as shockwave area, shockwave speed, and queue length, serve as temporal and spatial representations of the environment space. These allow for the approximation of relative positions of individual vehicles at the cycle level as well as the total stopped time. Given the relationship between traffic stability and shockwave parameters, more complex relationships between conflict rates and signal-vehicle optimization may be learned when a reinforcement learning approach is applied.

The signal state may be represented as a binary value, indicating red or green for each approach. However, this may result in issues for the reinforcement learning algorithm due to the

potential for the data to be treated as a non-binary values during backpropagation calculations. As such, an alternative measure was devised to allow the state to be represented by considering the remaining green and remaining red time on each approach. This would allow for a temporal representation of both trajectory optimization and signal timing. “Remaining Green Time” would later be compared against “Time to Intersection” to simplify the trajectory optimization problem.

When considering the volume, platoon ratio, and shockwave parameters, a near-infinite state space can be discretized such that the traffic conditions on a given roadway for the purpose of reinforcement learning can be approximated using a single score. Due to limitations with respect to the number of connected and conventional vehicles on the road, these parameters must only consider connected vehicles, and non-connected vehicles are to be ignored for the purpose of state representation due to the signal sensors' inability to properly detect conventional vehicles. For the purpose of state representation, conventional vehicles are assumed to be effectively invisible.

Lastly, the timesteps used are an important consideration in representing the environment state and should be appropriately selected. Any reinforcement learning agent used will decide upon the best possible action following a given timestep using the knowledge available at a given point in time. For the purpose of trajectory optimization alone, it is possible to devise a system that would ignore cycle-level parameters and work using instantaneous data. When considering signal timing, any timestep with a duration less than at least the current phase would rely on rolling averages and multiple decisions prior to a proper signal extension. To circumvent this issue, each timestep was assumed to be equal to the time it took to complete a phase including the extension time. Thus, all actions are to be performed either at the start of a new phase or at the end of the current extended phase.

4.1.3 Action Representation

An SVCC system should be designed to control both the vehicle trajectories as well as the signal timing. With regards to trajectory optimization, multiple advisories are required to accommodate the environmental constraints involving multiple approaches with multiple lanes. Ideally, each vehicle should receive its own individual speed advisory. Unfortunately, doing so would result in large complexity, hindering convergence. An alternative approach would be to consider the time it takes for a vehicle to cross the 225m span and designate a target speed to all vehicles crossing this point at a given point in time, and categorize them as belonging to part of a platoon. Vehicles that are not expected to arrive in time may be considered as part of a different platoon than those that arrive in time. This is represented by a variable known as "time to signal", which estimates the expected arrival time and compares it to the remaining signal time on the approach's current green or red phase.

The action space can be further simplified by considering simple solutions to the problem. Namely, vehicles that may pass through green should not slow down, and that the vehicles that cannot should slow down. This approach is similar to that taken in Chapter 3. With this logic in mind, two slowdown speeds must be identified, representing the two approaches where a slowdown is warranted at a given phase. The slowdown speeds selected are applied to platoons based on a trajectory projection given the current speed and position of entering vehicles. Applying these speeds at the sensor range (225 m) maximizes the time it a vehicle can take to adhere to the speed advisory, improving performance. As in Chapter 3, the selected speeds were assumed to follow a speed distribution similar to that of a speed limit, but with a shifted mean.

With regards to signal optimization, an additional action may be represented as the ideal phase extension time such that conflicts are reduced. Ranging from 0 to 20 seconds, the current green time of any given phase may be extended, followed by an amber phase, an all-red phase,

and an adjacent green phase. This continuous representation would allow for a finer calibration of extension times every phase. A single timestep would thus be given by the following, where $Green_{extend}$ represents the range of [0,20] and $Green_{min}$ represents the minimum green time of the adjacent approach.

$$Green_{extend} + Amber + All\ Red + Green_{min}$$

4.1.4 Reward Representation

Any reinforcement learning approach requires the selection of an appropriate reward function to quantify both short and long-term benefits. Common issues with RL approaches include the possibility of optimizing for and arriving at a local minimum, as well as optimizing for one component in the system at the expense of another if the reward function is not appropriate. Thus, a holistic approach is required to be effective at optimizing the entire system. In the context of developing an SVCC, this translates to optimizing all four approaches simultaneously given that in many cases an improvement on one approach may result in a worsening of traffic conditions on another.

$$\boldsymbol{E}(\boldsymbol{Y}) = -\sum_{App=1}^{4}\sum_{Lane=1}^{Lane_n} V^{1.239} exp(-1.624 + 0.294\,A - 0.828\,P + 0.119\,S_{12})$$

A reward function was thus selected to incorporate the safety performance functions derived by Essa et al. 2018 (Essa & Sayed, 2018), and define performance as the summation of conflicts on all lanes and all approaches. A negative sign is used to indicate that a decrease in conflicts is a positive action. The rewards are to be computed every timestep and changes in the reward can be easily identified by the system over time, allowing the agent to modify its policy.

4.1.5 Agent Selection

A Soft Actor-Critic (SAC) Agent was selected as a reinforcement learning agent to optimize for safety. Unlike the DDPG agent used in Chapter 3, SAC agents are stochastic in nature

and better suited towards evading local maxima to find superior solutions to complex problems. The selected agent was created using the MATLAB Reinforcement Learning Library, and featured a critic network and an actor-network with two hidden layers of 400 and 300 neurons. A continuous action space was represented by the outputs, the first of which represented the signal extension time ranging from [0,20], and the second and third representing the ideal slowdown advisory speeds ranging from [20,50], assuming a speed limit of 50 km/hr. The simplified action space was used to reduce the computational intensity and ensure quicker convergence.

To account for the rewards of subsequent actions, a look-ahead factor of 5 was used in the computation of future rewards to establish a policy. This value was selected such that the time period considered at each timestep corresponds to a larger time than the cycle time. This allows the discounted cumulative value of actions determined by the critic to be computed on a cycle level, rather than only a single phase, more accurately representing long-term rewards. The look-ahead factor of 5 was thus selected after minor iteration. Relationships between subsequent observations, namely relating to the signal state, were used to accomplish this goal.

4.1.6 Speed Assignment

Platoon behavior is an important component that can influence the effectiveness of trajectory optimization techniques. GLOSA techniques used typically provide individual users with discrete ideal speeds or a range of speeds to optimize throughput and minimize delays. These are typically performed once per signal based on the sensor range. Due to the scope of this study only including one intersection and a span of 225 m, vehicles beyond this point may not receive speed advisories and the system is unaware of their properties. Commands may only be sent once the vehicle or platoon is in range.

In theory, an ideal arrival pattern at the intersection stoplight would be defined by regular and predictable vehicle flows arriving during the green phase. Thus, the speed assignment system

must attempt to replicate these patterns by organizing vehicles into platoons on the approach preceding the intersection. To maximize the available time for this to occur, a series of rules should be established.

Firstly, the process involves identifying all incoming vehicles that enter the sensor range 225 m from the intersection. If the vehicles have connective capabilities, their positions, speeds, and accelerations are broadcasted to the controller. The controller then compares the expected arrival time at the signal stop line via trajectory prediction to the remaining signal time on the green phase to create a filter. The vehicles that are able to pass are not given any additional instructions whereas the vehicles that are unable to pass would be informed ahead of time to slow down to an acceptable speed which would inevitably result in the formation of a new platoon, splitting the existing one. The signal to slow down is sent based on the remaining signal time per timestep and applied at the sensor range to maximize action time. Vehicles already within sensor range prior to the order that are unable to clear the intersection will receive a slowdown order to help dissipate shockwaves. This may result in platoons splitting up into two parts; those scheduled to cross the intersection and those told to slow down.

The slowdown speeds that are selected remain active until the next green phase. This behavior ensures that incoming vehicles at a red light will slow down appropriately minimizing wait times. A speed-up order is sent to all stopped vehicles prior to the green light to minimize start-up delay and ensure situational awareness for the vehicles in these platoons. Vehicles expected to arrive on a green light are also to be told to speed up accordingly.

4.1.7 Process Overview

The proposed system consists of a central signal controller with the ability to receive and transmit data relating to vehicle speed and a position with an assumed range of 225 m. At the end of every phase, the signal controller would begin the process by receiving transmitted data from

each of the connected vehicles in the area. The data relating to speed and position would then be used to predict vehicle trajectories. This information is then used to predict the expected current conflict rate which is used as an input to the reinforcement learning algorithm.

The signal controller may then extend the signal time by a certain amount and issue speed advisories to incoming vehicles simultaneously. This is possible since there is only one RL agent which is used to decide upon multiple actions. Thus, the signal controller would be able to issue speed advisories to platoons of entering vehicles, requesting that they accelerate/decelerate safely to the ideal target speed to clear the intersection well within the extended signal time.

On approaches where the current signal state is red, stopped vehicles would be alerted ahead of time of the time remaining until the next phase, resulting in improvements in reaction time and potentially reducing perceived wait times.

4.2 Validation and Testing

4.2.1 Modelling and Representing Vehicle Behaviour

The proposed algorithm was modeled on a simulated intersection in order to evaluate the performance of the system in the field. PTV VISSIM simulation software was used to represent the intersection environment, and the Wiedemann 99 car-following model was used to represent vehicle's behavior. Using the simulation software, vehicle behavior was simulated in real-time and advisories were issued via a COM interface.

Test intersections were created and calibrated based on real-world data obtained from two intersections in Surrey, BC, 128th Street and 72nd Avenue and 132nd Street and 72nd Avenue, the same intersections used in the previous chapter. The signal timing plans for each intersection were obtained and applied during each cycle using an external COM interface. Existing detector loop systems were also modeled, allowing the signal timing plan to be adjusted for conditional left-turn phases.

Much like in Chapter 3, Connected vehicles were modeled as a special vehicle class that has a unique ID that can be detected by the controller system. This prevents data from conventional vehicles that would not be available in the field from being used by the controller in deciding signal timing. I2V and V2I latency for these vehicles was assumed to have a negligible effect on the final results and therefore were not considered for the purpose of this study.

Connected vehicle data were used by the controller to decide the actions to apply in each timestep. The signal time extension is applied to the signal groups in VISSIM by modifying the existing signal timing plans by the amount decided upon by the RL algorithm. Dynamic speed advisories used for trajectory optimization were applied using desired speed decisions. When a vehicle crosses a desired speed decision point, it is assigned a desired speed that follows a given speed distribution similar to a speed limit. This results in a variance in the acceleration and deceleration behavior observed and varying degrees of compliance with the speed advisory.

4.2.2 Defining the Test Sites

A total of two intersections were modeled with varying scenarios to represent different arrival patterns and traffic volumes throughout the day. The same test sites used in Chapter 3 to develop the proposed trajectory optimization system were used to test the SVCC algorithm. The first intersection at 128th Street and 72nd Avenue was used to train the reinforcement learning model while the second intersection was used to validate the results.

While the characteristics of the two intersections may differ, it is important to observe the effect of the proposed system on a wide variety of different approach types to gauge effectiveness. Should the system perform similarly on both intersections with a net safety improvement prior to additional training, it could be said that the system is transferable and effective. Additional analyses with respect to the effect of training can be observed to determine the benefits of operating the system over a longer period of time.

The hourly traffic volumes observed on 128th Street and 72nd Avenue ranged from 2,343 vehicles from 10:00 AM - 11:00 AM to 4,186 vehicles from 3:00 PM - 4:00 PM. The hourly traffic volume observed on 132nd Street and 72nd Avenue ranged from 2,140 vehicles at 10:00 AM - 11:00 AM to 3,354 vehicles at 5:00 PM - 6:00 PM. It should be noted that the southbound and northbound traffic volumes are for a single lane of traffic which differs from the two-lane northbound and southbound approaches on 128th Street.

4.2.3 Training the Model

The MATLAB Reinforcement Learning Library was used to implement the SAC model network. A series of scenarios were created for several hours of the day between 8:00 AM and 5:00 PM based on data collected at 128th Avenue and 72nd Street. 20 episodes of training were performed for each of the 9 simulation hours for a total of 180 episodes. This was followed by additional training sampled from random hours to ensure convergence to yield diminishing marginal improvements. In total, the system was trained for 500 episodes.

Reinforcement learning systems run the risk of generating an overfit solution. This was addressed by adding inherent randomness and varying the training times in each of the scenarios. Each episode consisted of a randomized sample of vehicles based on real-world traffic conditions at a given hour. To prevent the system from learning how to only optimize for a particular set of scenarios, random number generator seeds were applied to VISSIM as described in Chapter 3. The seeds allowed for some variation in vehicle generation and thus allowed for the system to learn patterns that were not specific to the scenarios used. The training was initially performed using several sets of the same scenario to generate a general solution. This was then adjusted to randomly sample from one of the nine scenarios available, each representing an hour of the day based on real-world data. This prevents an overfit and allows the controller to apply to situations with different volumes and vehicle arrival patterns.

To evaluate the average benefit expected from using this model, all vehicles on the road were assumed to be CVs as reflected in the initial training. Further testing examined the impact of conventional vehicles on the model's effectiveness.

4.2.4 Testing Procedure

The main objective of this algorithm is to substantially reduce the number of conflicts relative to the baseline model. A baseline, where all vehicles on the road were conventional, was created for each of the scenarios. The primary performance metric used was the conflict rate per simulation hour, which has a strong correlation to average delay. Thus, the effects of delay were also observed.

After training, an additional 90 episodes were run to evaluate the average conflict reduction in each of the nine scenarios. This allowed for the estimation of the total safety benefit for the intersection as well as each individual approach to ensure a holistic solution was found.

To ensure transferability, the trained model was applied to a secondary intersection (132nd Street and 72nd Avenue) to validate the results. Given the different physical characteristics of this intersection as compared to the first (128th Street and 72nd Avenue), it was expected that 132nd Street and 72nd Avenue would exhibit different behavior with regards to trajectory optimization. As a result, individual approaches were also considered in the analysis to better observe the differences in performance between them.

Due to the complexities of mixed environments, the algorithm should be designed to function in the transitional period where not all vehicles on the road are CVs. Thus, an additional series of tests were performed on the trained model by varying the Market Penetration Rate (MPR) and the percentage of CV vehicles on the road.

4.3 Results

4.3.1 Observed Conflict and Delay Reduction

After testing the effects of the algorithm on both 128th Street and 72nd Avenue, as well as 132nd Street and 72nd Avenue, a total conflict reduction of up to 55% and 41% were observed respectively as shown in Table 1. This corresponds to an average conflict reduction from approximately 0.16 conflicts/hr to 0.07 conflicts/hr in the test intersection. These were accompanied by delay reductions of 24% 128th Street and 21% for 132nd Street. Due to the nature of reinforcement learning algorithms, additional training would increase the maximum over time and bring the average closer to this maximum value. These results represent the sum of all traffic conflicts relative to the baseline between the hours of 9:00 AM and 5:00 PM.

Table 4: Delay and Conflict Reduction at the Test and Validation Intersections

	128th Street and 72nd Avenue (Test Intersection)	**132nd Street and 72nd Avenue (Validation Intersection)**
Conflict Reduction	55%	41%
Delay Reduction	24%	21%

Following convergence, the results showed a statistically significant reduction in traffic conflicts with a standard deviation of 7.7% and 4.5% for 128th Street and 132nd Street respectively. As such, at 100% MPR, the system performs well and can be expected to halve the number of rear-end collisions most of the time. System performance throughout the day varies and is most effective in time periods with higher traffic volumes. At lower volumes, trajectory optimization governs due to less obstruction and simpler conditions, whereas at higher volumes, the traffic signal optimization governs due to its ability to substantially reduce waiting times at the intersection. The reduced stopped times influence several parameters correlated to rear-end conflicts, resulting in the behavior observed in Figures 12 and 13.

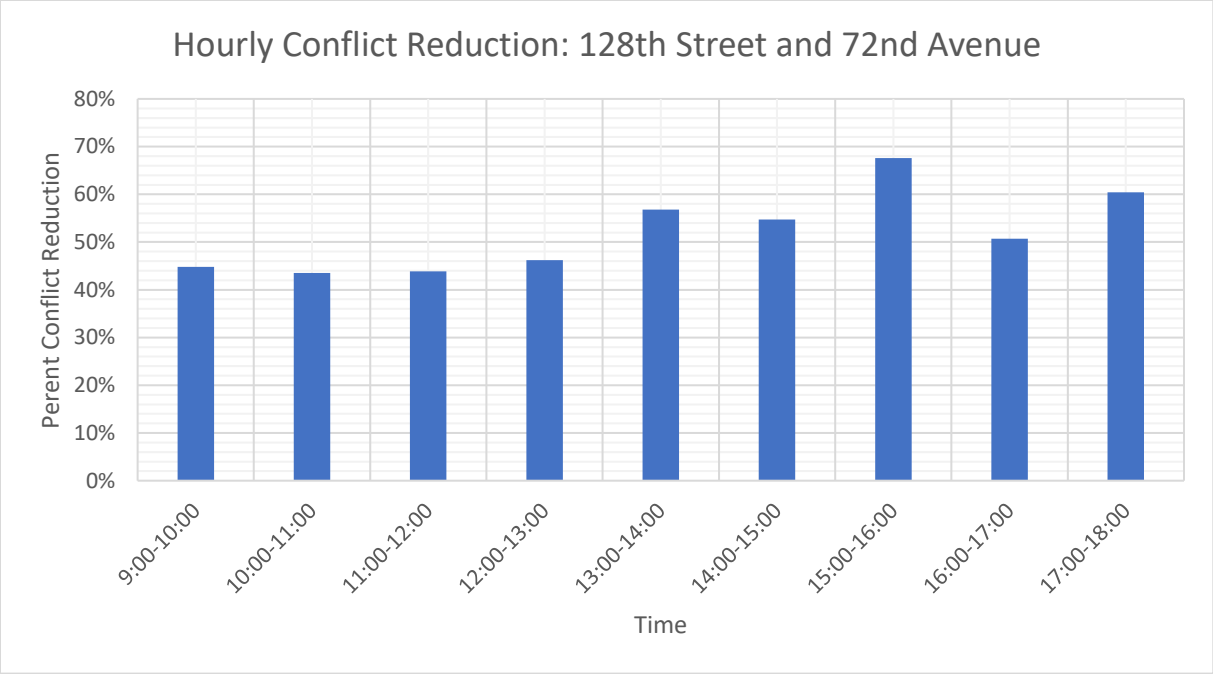

Figure 12: Hourly Conflict Reduction: 128th Street and 72nd Avenue

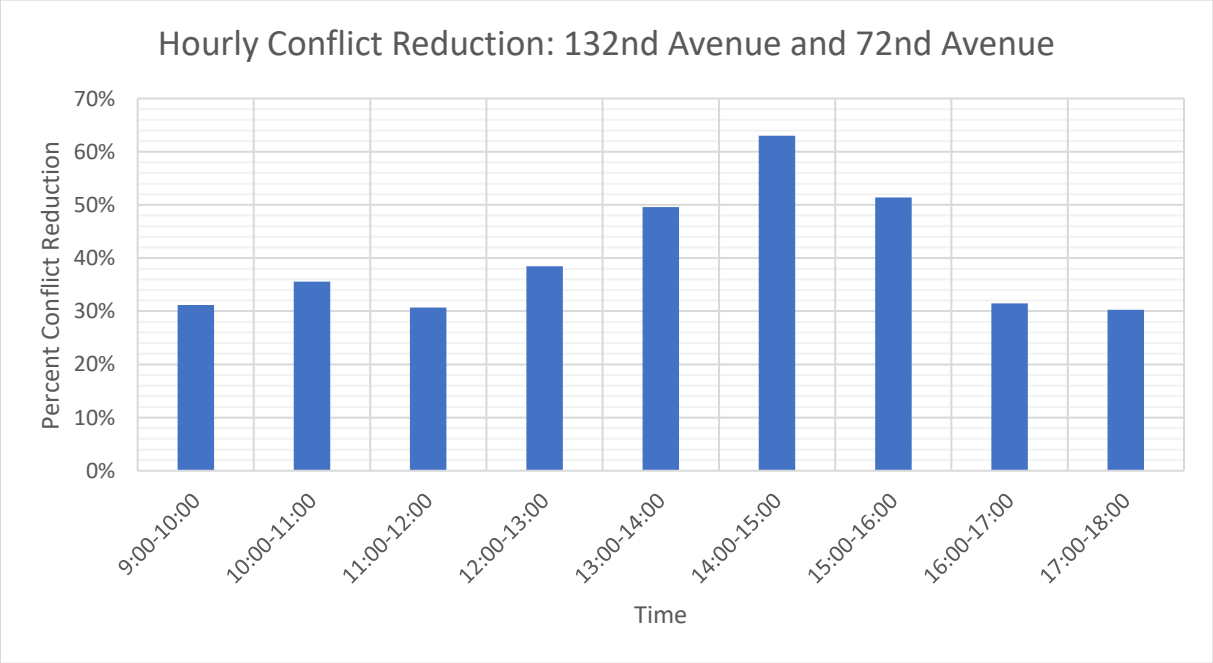

Figure 13: Hourly Conflict Reduction: 132nd Street and 72nd Avenue

To ensure that all approaches are well optimized, the data were further organized into approaches to visualize the effect of the algorithm. Overall, the results on 128th Street and 72nd Avenue are reasonable and show a significant conflict reduction on all approaches. On 132nd Street and 72nd Avenue, the north-south approach was improved significantly more than the east-west approach. This is a result of existing conditions on the one-lane roads being dramatically changed by the signal time extension, leading to large reductions in stopped time and thus, conflict reductions. This situation is seen in Figure 14, primarily between the hours of 13:00 and 16:00 where the percent reduction in conflicts on the east-west direction is significantly less than that on the north-south direction. This nonetheless nets a significant reduction in conflicts.

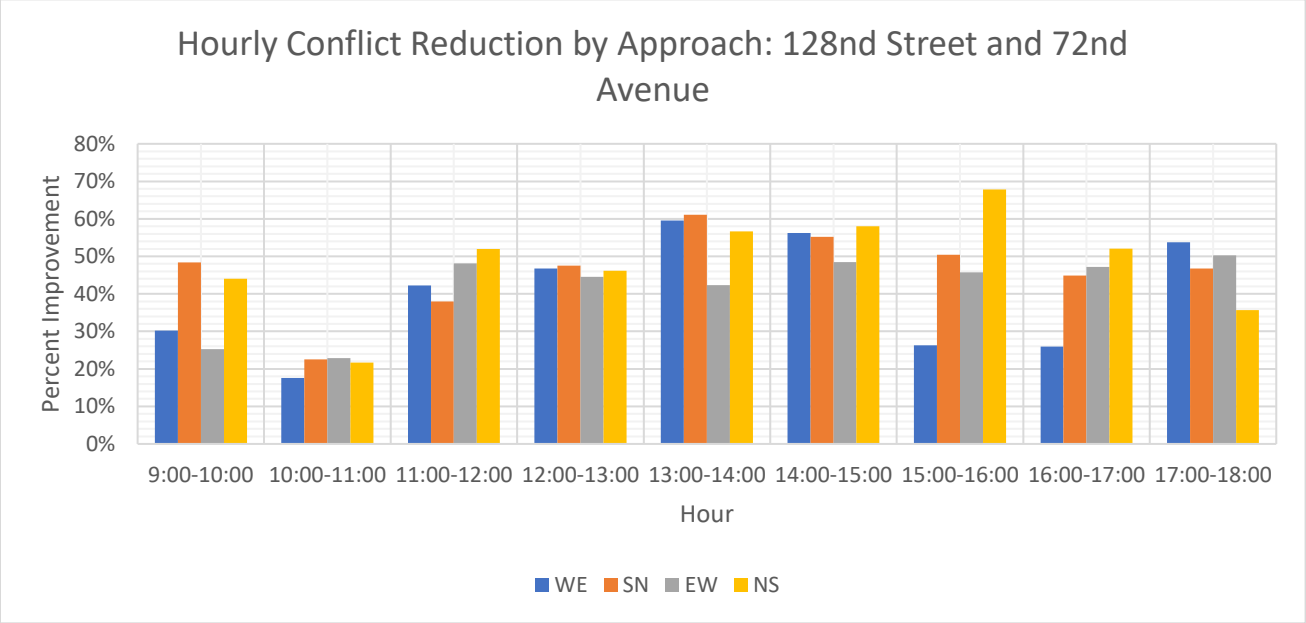

Figure 14: Hourly Conflict Reduction by Approach – 128th Street and 72nd Avenue

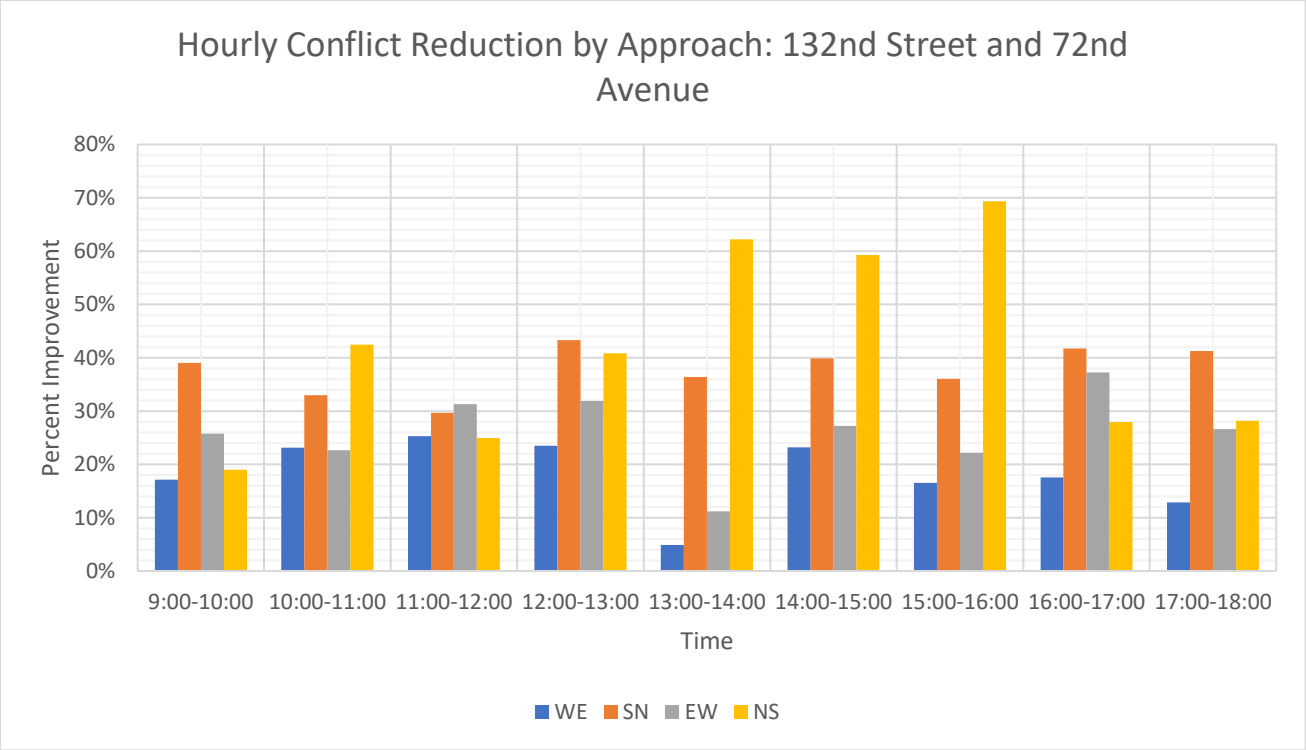

Figure 15: Hourly Conflict Reduction by Approach – 132nd Street and 72nd Avenue

4.3.2 Effect of MPR in the Transitional State

To visualize the performance of the algorithm in the transitional state, where not every vehicle has connected capabilities, several tests were conducted to estimate performance at various levels of market penetration. At the lowest level tested (25%), there were no observed adverse effects on safety or delay despite 75% of vehicles going undetected by the system as shown in Table 5. This indicates that the presence of a few individual CVs transmitting their positions and speeds can act as a representative sample for a large number of conventional vehicles.

For the intersection at 132nd Street and 72nd Avenue, an MPR of 25% results in a 31% reduction in rear-end conflicts, representing over half of the total possible conflict reduction obtained with this algorithm at 100% MPR. This demonstrates that nearly half of the safety benefit can be obtained very early in the implementation process. Similarly, at 128th Street and 72nd

Avenue, an MPR of 50% yields a total conflict reduction of 45%, which is almost 90% of the total safety improvement.

Table 5: Effect of Varying MPR on % Conflict Reduction

MPR:	25%	50%	75%	100%
% Conflict Reduction 132nd Street and 72nd Avenue	31%	39%	41%	41%
% Conflict Reduction 128th Street and 72nd Avenue	43%	45%	48%	55%

4.4 Discussion

4.4.1 Observed Vehicle Behaviour

To observe the performance of the system, the position of each vehicle was recorded over time and plotted into a space-time diagram. The situation before and after the implementation of the system was plotted, showing significant reductions in the shockwave area. Figure 16 illustrates the situation and highlights the shockwave areas in red. Queue length, shockwave area, and stopped time were significantly reduced by the proposed system. This can be attributed to the signal optimization portion of the algorithm as well as the trajectory optimization component.

The trajectory optimization portion of the algorithm can be observed in the diagram as shockwaves are dampened vertically by slowed vehicles ahead of a red light. This results in a reduction in the slope of the connected vehicles on the space-time diagram, relative to the conventional vehicle scenario. The signal optimization component, however, worked to reduce the width of the shockwave areas. The width represents stopped time and was reduced using variable signal timing. While Figure 16 demonstrates one example of system performance, it should be

noted that performance varies depending on the time of day, the approach, and the subject intersection.

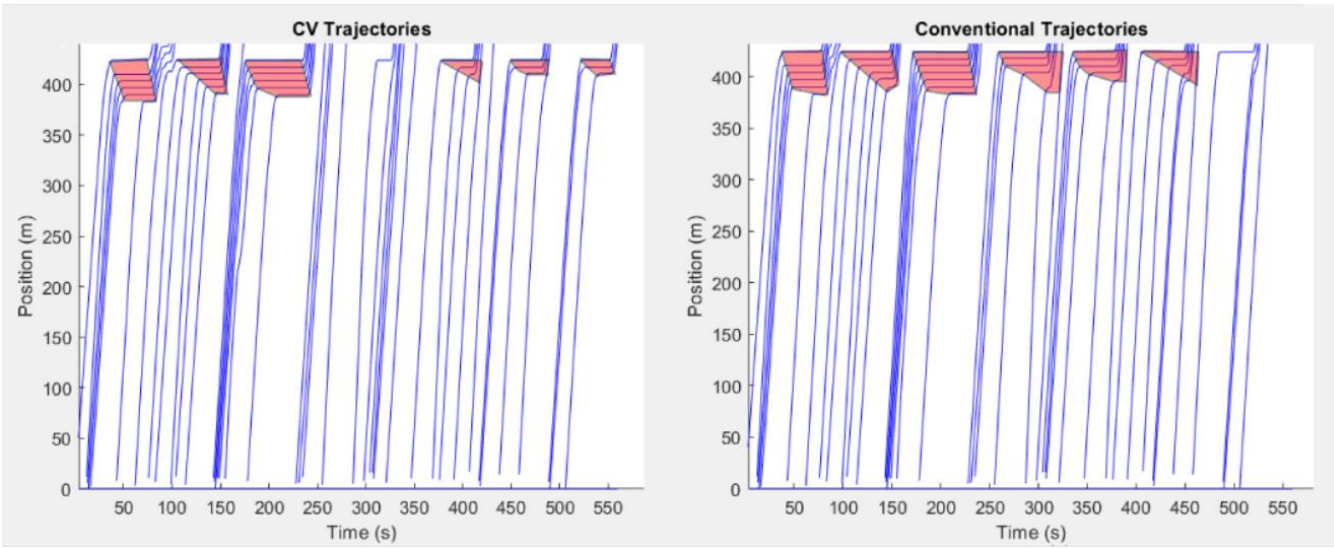

Figure 16: Sample Space-Time Diagram of NB 132nd Street and 72nd Avenue

4.4.2 MPR Analysis

The MPR analysis indicated that much of the benefit can be obtained prior to full CV implementation. To better explain this behavior, the VISSIM simulation runs were observed during the low MPR scenarios and the trajectories of individual vehicles were recorded over time.

On 132nd Avenue, a small number of connected vehicles permit the signal controller to modify the timing plan to allow for significantly more vehicles in the north-south direction to cross when necessary. Trajectory optimization helped dampen the size of the shockwaves, but the primary mechanism by which the conflicts were reduced was the decrease in north-south stopped time through signal optimization. Unfortunately, much like the scenario described in the previous Chapter, the presence of a single through and right lane may have adversely affected the trajectory optimization algorithm. The controller occasionally provided a leading vehicle with a “slow down” speed advisory during a red light, only for it to turn right once it reaches the intersection. The

system may have caused unnecessary queuing and delay if it was possible for this vehicle to turn earlier.

Furthermore, despite the lower percentage of CVs on the road, the ratio of CVs on the various approaches, on average, adequately reflects the traffic volumes in all directions, allowing the signal controller to decide upon effective actions.

On 128th Avenue, the presence of similar approaches on all four sides of the intersection played a role in assisting in trajectory and signal optimization alike, yielding a large conflict reduction at low levels of MPR, with much of the benefit obtained prior to 50% MPR. Due to the presence of multiple lanes on all approaches, the dampening effect of the speed advisories prior to a red light allowed for significant reductions in conflicts relative to a one-lane road where permissive turning movements are unnecessarily slowed down.

It is important to note that while the results indicate significant safety improvements at low levels of MPR, the actual results in practice are likely to vary significantly as there is a high degree of uncertainty associated with the transitional state. The effectiveness of trajectory optimization may be highly dependent on the random sample of vehicles and the distribution of connected vehicles within a given platoon.

4.4.3 Effect of Sensor Delay

The proposed system works under the assumption that the issuance of speed advisories and input of parameters are instantaneous. This assumption was made as the system timescales work at a phase level. Sensor delay operates at an order of magnitude of less than one second as data is transmitted from the CV to a roadside unit, processed, and sent back to the CVs. As such, had this study accounted for signal processing delay, the data used as part of the inputs and the reward function would simply be shifted backward by approximately one second. Preliminary tests were

performed to observe the effects of this delay by sending speed advisories to vehicles with a one-second delay. The results were not significantly affected.

4.4.4 Sensor Corruption and Resiliency

In addition to studying the effects of MPR on performance, this study indirectly examined the effect of erroneous data omissions resulting from faulty CV sensors. Since the algorithm does not consider conventional vehicles in its decision-making process, CVs transmitting faulty or no data would simply be ignored. At any given point in time, the signal controller would receive hundreds of signals, and as such, it is extremely unlikely that an objectively net-negative action would be taken in a high MPR scenario. At lower levels, it is still unlikely that a significant number of vehicles would be transmitting faulty data. Nonetheless, if the system is to be implemented, safeguards should be put in place to detect whether incoming data used for signal-vehicle control is consistent with previously transmitted data.

4.4.5 Impact of Intersection Characteristics

Although the sample size remains relatively small with only eight approaches being tested, there is some evidence that the benefits of the vehicle control portion of the algorithm are greater under certain conditions. The vehicle speed control system operates by issuing a multi-lane speed advisory for vehicles entering the intersection. These advisories would affect the speed of vehicles as they clear the intersection and have a slight effect on their start-up acceleration if they are stopped at an intersection. As such, vehicle platoons may act differently depending on the number of lanes and the ability to turn either left or right on a specific lane.

In two-lane road segments, the system appears to operate as expected without generating additional shockwaves. Speeds were appropriately assigned and strange lane changing behavior was not observed. Right-turning and left-turning vehicles were able to switch lanes with ease during a slowdown command and make appropriate movements.

However, in one lane approaches, right-turning traffic, which shares a lane with through traffic, saw significant issues. The algorithm does not receive signals from the vehicles informing the system of their intended direction. Rather, the algorithm only assumes their trajectory based on the current position, speed, and acceleration. As such, a right-turning vehicle on a single through-right lane would slow down during a red phase despite potentially being able to turn. This unnecessarily increases the queue length at the intersection while occasionally creating additional stop-and-go waves upstream. The negative effects of this aspect of the algorithm appear to be largely offset by the signal optimization algorithm which works to prevent these occurrences.

Chapter 5: Conclusion

This thesis proposed two algorithms that utilize real-time data obtained from V2I sensors on connected vehicles to improve safety at intersections. Hybrid dynamic-programming-reinforcement learning approaches were used to accomplish this goal primarily by seeking to reduce the incidence of rear-end conflicts.

The first algorithm used a DDPG Reinforcement Learning algorithm to minimize the shockwave area, assuming that it acts as an appropriate surrogate for overall safety. This assumption was then verified to be true. The continuous state space was discretized and split into key variables relating to occupancy of certain sections of the road, speed of vehicles occupying a certain roadway section, the gap between platoons, and signal timing parameters. Subsequently, these discrete inputs were considered by the algorithm which provided vehicles with a real-time speed advisory and returned a near-instantaneous real-time reward based on the shockwave area.

The second algorithm presented an SVCC system that optimizes safety in real-time using a hybrid dynamic programming-SAC reinforcement learning framework. Data from CVs were used to predict future vehicle trajectories and approximate the current conflict rate at the intersection. The proposed system uses a centralized signal controller to decide whether to extend the current phase based on the environment state, and by how much. The problems of signal optimization and vehicle trajectory optimization are simultaneously addressed by the proposed system. Vehicle trajectories are altered using speed advisories issued to platoons of incoming vehicles, with multiple ideal target platoon speeds decided upon using the reinforcement learning agent. Much like the first algorithm, the second algorithm discretizes a continuous state space and applies discrete actions to improve safety. The second algorithm differs from the first by using additional shockwave parameters to approximate both the current state and establish a reward.

Both algorithms were trained using PTV VISSIM simulation software, making use of its COM feature to send commands to modify individual vehicle's desired speed during the simulation. In the first system, a total of eight different speeds were decided upon every five seconds once the simple solutions of maintaining the same speed or accelerating to the speed limit were eliminated from the action space. The second system builds off the first and establishes a speed advisory system with larger timesteps, based on the sum of the remaining signal time and signal time extension, to decide whether a vehicle should accelerate to a certain speed or decelerate.

The simple speed advisory system yielded a notable daily conflict reduction of 23% on average, ranging from 2% to 37% on the first study intersection. On the second intersection, with different characteristics, a conflict reduction of 9% was observed with adverse effects on safety observed on the southbound single lane approach which did not have an exclusive right turn lane. The data was then filtered to ignore approaches with significantly different characteristics and a conflict reduction of 15% was observed. This compares to the more sophisticated SVCC system which yielded an average daily conflict reduction of 55% and 41%. These results were observed on each of the two intersections, indicating a statistically significant improvement in safety. Similarly, delay was tested and conflict reductions of 24% and 21% were observed. The system was further tested on individual scenarios representing a given hour between 9:00 AM and 5:00 PM. The results were filtered by approach to ensure that the system performs as designed. Conflict reductions ranged from 5% to 69% on an approach level and from 30% to 68% when considering the intersection as a whole. As such, it can be said that both systems present a holistic approach and are not expected to adversely affect one approach in improving another.

Recognizing the need for testing during the transitional state where not all vehicles on the road may transmit and receive data, both systems were tested on varying MPRs to determine their respective effectiveness. The speed advisory system demonstrated that the majority of the benefits can be obtained at relatively low MPRs such as at 25%. This, however, results in a greater variance in benefits depending on where the controlled vehicle is in a given platoon. A similar effect was observed in the SVCC system in which the majority of benefits were also obtained at low MPRs, indicating tangible benefits prior to full implementation. The feasibility of intelligent CV-based signal control and speed advisory systems has potential policy implications in the near future as a result.

Despite the effectiveness of these CV-based systems, several limitations may affect performance in the field. As with any road safety interention, additional case-specific studies should be performed before implementation. This study examined the effects of the proposed algorithm on model intersections with four approaches. Since the two intersections differ in configuration, different results were obtained. In the first system, changing the number of lanes appeared to affect performance. The presence of right and left-turn lanes also influenced system performance. As such, while an inference can be made for future projects, intersections with different characteristics may see different effects should this framework be implemented.

Furthermore, the impact of erroneous data should be considered. At low levels, faulty speed and position readings would only slightly affect the results with a near negligible effect similar to a change in MPR. However, over time, unless there is a legislative push to ensure adequate maintenance and functionality of connected vehicle sensors, a trend resulting in large amounts of faulty sensors among the CV fleet would result in poorer system performance. Sensors transmitting faulty data are a greater risk than sensors transmitting no data. A built-in safeguard for this scenario

is the other variables that are considered, which may prove to be appropriate predictors for the model. For example, if the number of vehicles in one quadrant is erroneously underreported, the speed of vehicles or the gap between platoons may overcome the inaccuracies and simply lead to a sub-optimal decision rather than a poor decision. Despite the benefits likely to be observed by implementing this algorithm, one should recognize that all systems have flaws and additional work should be performed in creating redundancies, especially when dealing with traffic safety and human drivers.

Another major concern is the assumption of a largely compliant CV fleet that abides by speed advisories. In the context of the model proposed in this study, non-compliant connected vehicles are modelled as conventional vehicles. In practice, extensive education campaigns will be required to explain to road users that slowing down when requested may reduce congestion. As noted by Matowiki et al. 2016 (Michał & Ondřej, 2016) in the trial of variable speed limits, human drivers could potentially opt to ignore the speed advisories in favor of a perceived personal benefit rather than a system optimum solution, potentially endangering themselves or other drivers. Additionally, the system assumes that certain "pacer vehicles" that are connected may work to slow down and regulate the speed of conventional vehicles behind them. These vehicles may face pressure from nearby motorists to speed up despite the warnings provided that they will not arrive at the intersection in time. As a result, dangerous lane-changing behavior to pass the connected vehicles may be observed. Thus, the implementation of the system would likely require some degree of public education, recognizing the limitations in the efficacy of said measures. Drivers will need to understand that slowing down and obeying the recommendations will result in shorter wait times and a safer drive. This benefit must be tangible and easily recognized to encourage motorists to comply. Any implementation of these types of systems would require additional

testing in the field to observe motorists' response to the advisories, and potentially improved enforcement mechanisms or incentives.

Reaction time delays due to input lag is another potential source of error not considered in the studies. It may be addressed by summing the average vehicle reaction time and sensor lag time to predict a vehicle's future position using basic kinematic equations. This updated position would "offset" the model by a few seconds and extend the refresh rate, allowing the system to be more precise. The final results would be slightly altered, and system resilience should be re-examined in future studies. According to the literature, delays in AV and CAV environments may result in significant traffic instability. (Yao et al., 2021) However, given that this thesis evaluates a CV environment without actuated control by the controller, significant instability resulting from sensor delays is unlikely to be observed.

Lastly, the devised algorithms worked to reduce rear-end collisions and did not consider other conflict types such as side-swipe conflicts or crossing conflicts. These conflicts were assumed to be constant and unaffected. In future studies, this assumption should be verified, and a more sophisticated algorithm should be devised to consider alternative conflict types. Further research should seek to incorporate additional data from CVs and adjacent intersections regarding long-term routing to help improve overall network performance. Additionally, there is great potential for the use of other roadside sensors to increase the range of the speed advisories along the approaches. A similar system can be devised for highways or longer road spans which would work towards magnifying the overall effect when observed on a network level. Ultimately, field testing of CV-based SVCC systems is required to fully validate the results and ensure that the theoretical benefits are observed in practice.

Bibliography

Ali, Y., Haque, Md. M., Zheng, Z., & Bliemer, M. C. J. (2021). Stop or go decisions at the onset of yellow light in a connected environment: A hybrid approach of decision tree and panel mixed logit model. *Analytic Methods in Accident Research*, *31*, 100165. https://doi.org/10.1016/j.amar.2021.100165

Ali, Y., Sharma, A., Haque, Md. M., Zheng, Z., & Saifuzzaman, M. (2020). The impact of the connected environment on driving behavior and safety: A driving simulator study. *Accident Analysis & Prevention*, *144*, 105643. https://doi.org/10.1016/j.aap.2020.105643

Arel, I., Liu, C., Urbanik, T., & Kohls, A. G. (2010). Reinforcement learning-based multi-agent system for network traffic signal control. *IET Intelligent Transport Systems*, *4*(2), 128. https://doi.org/10.1049/iet-its.2009.0070

Baibing, L. (2017). Stochastic modeling for vehicle platoons (I): Dynamic grouping behavior and online platoon recognition. *Transportation Research Part B: Methodological*, *95*(1), 364–377.

Beak, B., Head, K. L., & Feng, Y. (2017). Adaptive Coordination Based on Connected Vehicle Technology. *Transportation Research Record: Journal of the Transportation Research Board*, *2619*(1), 1–12. https://doi.org/10.3141/2619-01

Behl, M., & Mangharam, R. (2010). *Pacer cars: Real-time traffic shockwave suppression*. Real-Time Systems Symposium(RTSS), San Diego.

Blincoe, L., Miller, T. R., Zaloshnja, E., & Lawrence, B. A. (2015). *The Economic and Societal Impact Of Motor Vehicle Crashes, 2010* (NHTSA Technical Report DOT HS 812 013). National Highway Traffic Safety Administration. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812013

Camponogara, E., & Kraus, W. (2003). Distributed Learning Agents in Urban Traffic Control. In F. M. Pires & S. Abreu (Eds.), *Progress in Artificial Intelligence* (Vol. 2902, pp. 324–335). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-24580-3_38

Comert, G., & Cetin, M. (2009). Queue Length Estimation from Probe Vehicle Location adn the Impacts of sample size. *Eur, J, Oper Res.*, 196–202.

Comert, G., & Cetin, M. (2011). Analytical evaluation of the error in queue length estimation at traffic signals from probe vehicle data. *Intelligent Transportation Systems*, *12*, 563–573.

Dresner, K., & Stone, P. (2004). *Multiagent traffic management: A reservation-based intersection control mechanism*. *2*.

El-Tantawy, S., Baher, A., & Hossam, A. (2013). Multi-Agent Reinforcement Learning for Integrated Network of Adaptive Traffic Signal Controllers(MARLIN-ATSC). *IEEE Conference on Intelligent Transportation Systems*, *14*.

Essa, M., & Sayed, T. (2018). Traffic conflict models to evaluate the safety of signalized intersections at the cycle level. *Transportation Research Part C*, *89*, 289–302.

Essa, M., & Sayed, T. (2020). Self learning adaptive traffic signal control for real-time safety optimization. *Accident Analysis & Prevention*, *146*.

Fazekas, A., Hennecke, F., Kalló, E., & Oeser, M. (2017). A Novel Surrogate Safety Indicator Based on Constant Initial Acceleration and Reaction Time Assumption. *Journal of Advanced Transportation*, *2017*.

Feng, Y., Head, K. L., Khoshmagham, S., & Zamanipour, M. (2015). A real-time adaptive signal control in a connected vehicle environment. *Transportation Research Part C: Emerging Technologies*, *55*, 460–473. https://doi.org/10.1016/j.trc.2015.01.007

Fink, J., Kwigizile, V., & Oh, J.-S. (2016). Quantifying the impact of adaptive traffic control systems on crash frequency and severity: Evidence from Oakland County, Michigan. *Journal of Safety Research*, *57*, 1–7. https://doi.org/10.1016/j.jsr.2016.01.001

Gaur, A., & Mirchandani, P. (2001). Method for Real-Time Recognition of Vehicle Platoons. *Transportation Research Record*.

Gu, J., Fang, Y., Sheng, Z., & Wen, P. (2020). Double Deep Q-Network with a Dual-Agent for Traffic Signal Control. *Applied Sciences*, *10*(5), 1622. https://doi.org/10.3390/app10051622

Guo, Q., Li, L., & Bana, X. (2019). Urban traffic signal control with connected and automated vehicles: A Survey. *Transportation Research Part C*.

Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018). Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. *ArXiv*.

Hao, P., Ban, X. (Jeff), & Whon Yu, J. (2015). Kinematic Equation-Based Vehicle Queue Location Estimation Method for Signalized Intersections Using Mobile Sensor Data. *Journal of Intelligent Transportation Systems*, *19*(3), 256–272. https://doi.org/10.1080/15472450.2013.857197

Hao, P., Sun, Z., Ban, X., Guo, D., & Ji, Q. (2013). Vehicle index estimation for signalized intersections using sample travel times. *Transportation Research Part C: Emerging Technologies*, *36*, 513–529.

Haque, Md. M., Oviedo-Trespalacios, O., Sharma, A., & Zheng, Z. (2021). Examining the driver-pedestrian interaction at pedestrian crossings in the connected environment: A Hazard-based duration modelling approach. *Transportation Research Part A: Policy and Practice*, *150*, 33–48. https://doi.org/10.1016/j.tra.2021.05.014

Hayward, J. (1972). Near-miss determination through use of a scale of danger,. *Highway Research Record*, *384*.

He, Q., Head, L., & Ding, J. (2012). PAMSCOD: Platoon-based arterial multi-modal signal control with online data. *Transportation Research Part C: Emerging Technologies*, *20*(1), 164–184.

Hiribarren G, H. J. C. (2014). Real time traffic states estimation on arterials based on trajectory data. *Transportation Research Part B: Methodology*, *69*, 19–30.

Hofleitner, A., Herring, R., & Bayen, A. (2012). Arterial Travel Time Forecast with Streaming Data: A Hybrid Approach of Flow Modelling and Machine Learning. *Transportation Research Part B: Methodology*, *46*, 1097–1122.

ICBC. (2020). *ICBC Statistics*. ICBC.

Iglesias, I., Isasi, L., Larburu, M., Martinez, V., & Molinete, B. (2008). I2V Communication Driving Assistance System: On-Board Traffic Light Assistant. *2008 IEEE 68th Vehicular Technology Conference*, 1–5. https://doi.org/10.1109/VETECF.2008.437

Ilgin Guler, S., Menendez, M., & Meier, L. (2014). Using connected vehicle technology to improve the efficiency of intersections. *Transportation Research Part C: Emerging Technologies*, *46*, 121–131. https://doi.org/10.1016/j.trc.2014.05.008

Islam, S. M. A. B. A., & Hajbabaie, A. (2017). Distributed coordinated signal timing optimization in connected transportation networks. *Transportation Research Part C: Emerging Technologies*, *80*, 272–285. https://doi.org/10.1016/j.trc.2017.04.017

Jiang, H., Jia, H., An, S., Wang, M., Park, B., & Byungkyu. (2017). Eco approaching at an isolated signalized intersection under partially connected and automated vehicles environment. *Transportation Research Part C: Emerging Technologies*, *79*, 290–307.

Kamonthep, T., Zhang, Y., & Ye, X. (2019). Platoon recognition using connected vehicle technology. *Journal of Intelligent Transportation Systems*, *23*(1), 12–27.

Karoui, M., Freitas, A., & Chalhoub, G. (2018). Efficiency of Speed Advisory Boundary fINder (SABIN) strategy for GLOSA using ITS-G5. *2018 IFIP/IEEE International Conference on Performance Evaluation and Modeling in Wired and Wireless Networks (PEMWN)*, 1–6. https://doi.org/10.23919/PEMWN.2018.8548912

Katsaros, K., Kernchen, R., Dianati, M., & Rieck, D. (2011). Performance study of a Green Light Optimized Speed Advisory (GLOSA) application using an integrated cooperative ITS simulation platform. *2011 7th International Wireless Communications and Mobile Computing Conference*, 918–923. https://doi.org/10.1109/IWCMC.2011.5982524

Kreideh, A., Wu, C., & Bayen, A. (2018). *Dissipating stop-and-go waves in closed and open networks via deep reinforcement learning*. 21st International Conference on Intelligent Transportation Systems (ITSC), Maui.

Li, W., & Ban, X. J. (2017). Traffic signal timing optimization in connected vehicles environment. *2017 IEEE Intelligent Vehicles Symposium (IV)*, 1330–1335. https://doi.org/10.1109/IVS.2017.7995896

Li, Z., Lily, E., & Sanjay, R. (2014). Signal control optimization for automated vehicles at isolated signalized intersections. *Transportation Research Part C: Emerging Technologies*, *49*, 1–18.

Lillicrap, T., Hunt, J., Pritzel, A., Heess, N., Erez, T., Yuval, T., Silver, D., & Wierstra, D. (2015). Continuous control with deep reinforcement learning. *ArXiv*.

Liu, W., Yin, Y., & Yang, H. (2015). Effectiveness of variable speed limits considering commuters' long-term response. *Transportation Research Part B: Methodological*, *81*, 498–519.

Lu Shoufeng, Ximin, L., & Shiqiang, D. (2008). Q-Learning for Adaptive Traffic Signal Control Based on Delay Minimization Strategy. *2008 IEEE International Conference on Networking, Sensing and Control*, 687–691. https://doi.org/10.1109/ICNSC.2008.4525304

Lu, X.-Y., & Shladover, S. E. (2014). Review of Variable Speed Limits and Advisories: Theory, Algorithms, and Practice. *Transportation Research Record: Journal of the Transportation Research Board*, *2423*(1), 15–23. https://doi.org/10.3141/2423-03

Ma, M., & Li, Z. (2021). A time-independent trajectory optimization approach for connected and autonomous vehicles under reservation-based intersection control. *Transportation Research Interdisciplinary Perspectives*, *9*.

Michał, M., & Ondřej, P. (2016). Speed Compliance in Freeway Variable Speed Limit System-Case Study of the Prague City Ring. *Transport Problems*, *11*(1).

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., I, A., D, W., & M, R. (2013). Playing Atari with Deep Reinforcement Learning. *ArXiv*.

Ozbay, K., Yang, H., Bartin, B., & Mudigonda, S. (2008). Derivation and validation of new simulation-based surrogate safety measure. *Ransportation Research Record: Journal of the Transportation Research Board*, *2083*, 105–113.

Pandit, K., Ghosal, D., Zhang, H. M., & Chuah, C.-N. (2013). Adaptive Traffic Signal Control With Vehicular Ad hoc Networks. *IEEE Transactions on Vehicular Technology*, *62*(4), 1459–1471. https://doi.org/10.1109/TVT.2013.2241460

Raj, K., Hesham, R., & Hao, Y. (2015). Networkwide Impacts of Vehicle Ecospeed Control in the Vicinity of Traffic Signalized Intersections. *Transportation Research Record Journal of the Transportation Research Board*, *2503*, 91–99.

Salkham, A., Cunningham, R., Garg, A., & Cahill, V. (2008). A Collaborative Reinforcement Learning Approach to Urban Traffic Control Optimization. *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 560–566. https://doi.org/10.1109/WIIAT.2008.88

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal Policy Optimization Algorithms. *ArXiv*.

Seredynski, M., Dorronsoro, B., & Khadraoui, D. (2013). Comparison of Green Light Optimal Speed Advisory approaches. *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, 2187–2192. https://doi.org/10.1109/ITSC.2013.6728552

Silver, D., Lever, G., Hees, N., Degris, T., Wierstra, D., & Riedmiller, M. (2014). *Deterministic Policy Gradient Algorithms*. International conference on machine learning.

Statistics Canada. (2020). *Leading causes of death, total population, by age group* [Data set]. Government of Canada. https://doi.org/10.25318/1310039401-ENG

Sun, W., Zheng, J., & Liu, H. X. (2017). A capacity maximization scheme for intersection management with automated vehicles. *Transportation Research Procedia*, *23*, 121–136. https://doi.org/10.1016/j.trpro.2017.05.008

Suramardhana, T. A., & Jeong, H.-Y. (2014). A driver-centric green light optimal speed advisory (DC-GLOSA) for improving road traffic congestion at urban intersections. *2014 IEEE Asia Pacific Conference on Wireless and Mobile*, 304–309. https://doi.org/10.1109/APWiMob.2014.6920310

Suzuki, H., & Marumo, Y. (2018). A New Approach to Green Light Optimal Speed Advisory (GLOSA) Systems for High-Density Traffic Flowe. *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 362–367. https://doi.org/10.1109/ITSC.2018.8569394

Tan, T., Chu, T., & Wang, J. (2020). Multi-Agent Bootstrapped Deep Q-Network for Large-Scale Traffic Signal Control. *2020 IEEE Conference on Control Technology and Applications (CCTA)*, 358–365. https://doi.org/10.1109/CCTA41146.2020.9206275

U.S Department of Transportation. (2015). *Connected Vehicles: Benefits, Roles, Outcomes*.

Wan, N., Vahidi, A., & Lucknow, A. (2016). Optimal speed advisory for connected vehicles in arterial roads and the impact on mixed traffic. *Transportation Research Part C: Emerging Technologies*, *69*, 548–563.

Wang, P., Wang, Y., Deng, H., Zhang, M., & Zhang, J. (2020). Multilane Spatiotemporal Trajectory Optimization Method (MSTTOM) for Connected Vehicles. *Urban Mobility Analytics Using Large-Scale Vehicle Trajectory Data*, *2020*.

Xie, X.-F., Barlow, G. J., Smith, S. F., & Rubinstein, Z. B. (2011). Platoon-based self-scheduling for real-time traffic signal control. *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 879–884. https://doi.org/10.1109/ITSC.2011.6082849

Xu, B., Ban, X. J., Bian, Y., Wang, J., & Li, K. (2017). V2I based cooperation between traffic signal and approaching automated vehicles. *IEEE Intelligent Vehicles Symposium*, *4*, 1658–1664.

Y Cheng, Qin, X., & Jin, J. (2013). An exploratory shockwave approach to estimating queue length using probe trajectories. *Transportation Research Part C*, *36*, 268–283.

Yao, Z., Xu, T., Jiang, Y., & Hu, R. (2021). Linear stability analysis of heterogeneous traffic flow considering degradations of connected automated vehicles and reaction time. *Physica A: Statistical Mechanics and Its Applications*, *561*, 125218. https://doi.org/10.1016/j.physa.2020.125218

Yingyan, L., Peiheng, L., & Xiaoyan, H. (2016). A distributed framework for network-wide traffic monitoring and platoon information aggregation using V2V communications. *Transportation Research Part C: Emerging Technologies*, *69*(1), 356–374.

Zhang, L., Zhang, L., Hale, D., Hu, J., & Huang, Z. (2017). Cycle-based variable speed limit methodology for improved freeway merging. *IET Intelligent Transport Systems*, *11*(10).

Zhang, X., Onieva, E., Perallos, A., Osaba, E., & Lee, V. C. S. (2014). Hierarchical fuzzy rule-based system optimized with genetic algorithms for short term traffic congestion prediction. *Transportation Research Part C: Emerging Technologies*, *43*, 127–142. https://doi.org/10.1016/j.trc.2014.02.013

Zheng, H. (2017). Estimating traffic volumes for signalized intersections using connected vehicle data. *Transportation Research Part C Emerging Technologies*, *79*, 347–362.