## Spillback Control at Intersections

**LIBREARE** 

In a Real-Time Controlled Traffic System with Multimodal Optimization

ū

Delft

C.F. Baak **Master Thesis** 





by



to obtain the degree of Master of Science

at the Delft University of Technology,

to be defended publicly on December 8th, 2022, at 15:00.

Dr.ir. H. Taale,

Student number: 4381629 Project duration: 27 March – 8 December 2022 Thesis committee: Prof.dr.ir S.P. Hoogendoorn, TU Delft

Dr. F. Schulte, TU Delft T. De Groot MSc, Technolution H.E. Mein MSc, Technolution

*This thesis is confidential and cannot be made public until 8-12-2022*

An electronic version of this thesis is available at [https://repository.tudelft.nl/.](https://repository.tudelft.nl/)



## Preface

<span id="page-4-0"></span>When I started my Master's in Transport, Infrastructure, and Logistics, I started doing the specialization 'operations', which focuses on the operational management and control of traffic and the technologies and methodologies facilitating it. It confirmed my interest in operational management and control, and after following the courses, I felt interested in graduating in this domain. I have been working on this thesis for the past eight to nine months. I am happy that I chose this research direction and that I learned about the complexity of an actual transport model, in which I was given an opportunity to design an additional component for. I think I will never look the same again at spillback effects when I am experiencing them myself or when I see them unfolding.

I did this thesis at Technolution as part of an internship. At the company, I would like to thank my supervisors, Tom and Edwin, for their input, availability and help in familiarizing myself with MobiMaestro-Flow. I enjoyed my time a lot at Technolution. I would also like to thank my graduation committee, which helped improve my work with their feedback. A special thanks to my daily supervisor, Henk Taale, who gave me very helpful feedback during the process and also suggested doing weekly syncs together with Tom and Edwin. This has helped me to maintain focus and motivation, but it also was a nice break from working on the thesis.

Lastly, I would also like to thank my parents and Eva for their endless support during this whole process.

*Cornel Baak Delft, 1-8*

## Signature page

<span id="page-6-0"></span>The following persons have approved the thesis as a complete and final work:

Prof.dr.ir. S.P. Hoogendoorn Dr. H. Taale Distinguished Professor Assistant professor TU Delft TU Delft

Dr. F. Schulte T. De Groot MSc TU Delft Technolution

Assistant professor Company supervisor

H.E. Mein MSc Company supervisor Technolution

## **Contents**













## List of Figures

<span id="page-14-0"></span>





## Summary

<span id="page-17-0"></span>Many configurations of traffic systems have been proposed, tested, and implemented to promote traffic flow and network performance in urban environments. When traffic demand is high and the intersection lacks throughput, traffic phenomena occur, such as oversaturation of a turning bay which leads to lane blockage. In addition, oversaturation of an intersection lane could lead to vehicles spilling over to an upstream intersection. These phenomena are called spillback effects and could lead to unsafe traffic situations. In some municipalities, traffic safety is assigned a high priority. In the United States, one-third of intersection fatalities still occur at signalized intersections, and existing real-time controlled traffic systems (RTCS) could contribute. By ensuring that emerging unsafe situations, such as spillback effects, are avoided or handled faster by providing additional priority in traffic light control. Priority in traffic light control is not new, as traffic light control with priority for transit already existed. Transit Signal Priority (TSP) has been used to improve the level of service of transit users and can be incorporated into the signal control. Existing signal plans are modified; however, the impact of multiple types of road users in urban areas increased the complexity of finding optimal signal control. Hence RTCS with multimodal optimization have been developed to optimize various road users in the same control objective. The control objective is either configured as the vehicle or personal delay. These multimodal RTCS have been tested in undersaturated conditions. However, they have yet to be refined to work in moderate or oversaturated conditions. Furthermore, research regarding multimodal RTCS that cope with oversaturated conditions (spillback effects) has been limited. Moreover, it is not clear what the consequences are of spillback control strategies in a multimodal RTCS. Although additional priority for spillback could be beneficial for avoiding or handling emerging unsafe situations. To fill in the knowledge gap a spillback component is designed and tested. More specifically, a spillback detection method and control strategy were designed to gain insight into the effects of applying spillback control strategies to traffic and network performance. This resulted in two main research questions: *How can an existing RTCS be extended with spillback detection and spillback control strategies in a multimodal context?*, and *What is the effect of controlling spillback in a multimodal RTCS at signalized intersections on the traffic and network performance?* To answer the research questions, a literature review was conducted to identify suitable approaches to detect and con-

trol spillback in an existing RTCS with multimodal optimization. Furthermore, a simulation study is performed to gain insight into the effects. The identified spillback detection and control methods were implemented as an extension to a multimodal RTCS. The utilized multimodal RTCS was MobiMaestro-Flow (MM-Flow). MM-Flow continuously optimizes individual intersections by minimizing the waiting costs over an optimization horizon of 120 seconds and an update frequency of 1 to 5 seconds. The waiting costs consist of the multiplications of the waiting time per vehicle, the movement, and object priority. The utilized algorithm is referred to as the Traffic Flow Engine (TFE). The literature review showed that the spillback detection method depends on the utilized traffic sensors. When using a more conventional sensor, such as a loop detector, mainly two approaches were used—first, detecting spillback by using a link capacity in terms of the number of vehicles, and second, by defining a queue threshold in meters. The approach with the queue threshold requires a queue method. In contrast, when more sophisticated sensors are used, spillback can be measured directly from video detection. Regarding the spillback control strategies, also two types were widely used. The first one solves spillback by an objective criterion in the optimization algorithm, promoting the flow of the congested link or the overall intersection flow. The other approach is metering. Vehicle inflow is limited upstream, the outflow is promoted downstream, or both are applied at the congested intersection.

Regarding the design of the spillback component, it was chosen to detect spillback by a queue threshold. The reason for this is that with a queue threshold (and a queuing model) spillback can also be seen in weather conditions (e.g. mist). Additionally, the required sensors are cheaper in purchase costs and the generated shockwave profile can also be used for delay estimation. The threshold is either the length of the link or the length of a turning bay. Spillback is detected when the estimated maximum queue length (*L n max*) during a traffic light cycle *n* exceeds the queue threshold. In terms of the spillback control strategy, two spillback control strategies were designed as an extension to the TFE [\(Figure 1\)](#page-18-0). The main difference between the two control strategies is that the first strategy alters the optimization outcome by enforcing green time for the congested direction. The second spillback control strategy penalizes the direction affected by spillback so that when spillback is detected, the direction receives higher priority in optimization. This leads to an earlier green and advantages for the congested link.



<span id="page-18-0"></span>

Figure 1: Implementation of control strategies in TFE structure

The effects of the spillback components were tested according to several experimental scenarios. The scenario were distinguished by the used simulation network model, the included modes, demand, and the spillback handling method. The experiments were performed at an isolated intersection and a corridor consisting of ten intersections comparable to a part of 3rd Street in San Francisco. For each experiment, data was collected for the spillback duration, the delays of six routes, and the network performance. The demand in the first quarter was halved, in the second and third quarter the demand factor was one, in the fourth quarter demand was halved again. In the experiments, the vehicles were prioritized as in [Table 1](#page-18-1)

The outcomes of the experiments showed that, in general, the intersection performance decreases when a spillback con-

<span id="page-18-1"></span>Table 1: Vehicle priority configuration



trol strategy is applied in traffic light control optimization. Regarding the height of the spillback penalty, only when the demand for disadvantaged directions is low enough that a higher spillback priority increases intersection performance. The provision of additional spillback priority causes unbalanced allocations of green time, meaning that in higher demands, the overall performance of an intersection is only increased without spillback control. In terms of spillback duration, a small spillback penalty leads to the best results when the demand is higher. A bus line, in combination with a low spillback penalty, improved spillback directions and decreased queue durations for the bus direction. With increased penalty factors, traffic control leans into an advantage for spillback directions causing higher delays for the bus and non-spillback directions. In a network, the intersection performance is most often increased at the ends of the corridor and the intersection downstream of the intersection affected by spillback. Regarding the network, the performance increased with a moderate vehicle flow and a minor spillback penalty. Furthermore, the network performance increased for all spillback control strategies during the peak period. Except when additional priority was provided for the arterial of the corridor.

The experiments at the corridor showed that a spillback priority effectively decreases spillabck durations and works well in combination with higher prioritized public transit. When a bus was included in the traffic flow, lower delays and spillback durations were found for spillback directions. However, the effect was paired with higher delays for disadvantaged directions; when buses travel in the spillback direction, spillback priority leads to practically absolute priority.

In terms of the height of the spillback penalty, a too-high priority negatively impacts the intersection and network performance. However, in some cases, a (high) spillback priority is necessary to provide enough priority so that a growing unsafe situation is avoided. Therefore, it is crucial to determine the intended use and applicability when assigning spillback priority. When an unsafe situation emerges, a very high spillback priority can be feasible, for example, when preemption is needed for an emergency vehicle and vehicles are spilt over at an intersection.

The designed methods and performed experiments are paired with limitations. First, regarding the developed control method, the proposed method cleans up instead of avoiding the spillback effect. Additionally, the control method assigns additional priority when a queue length exceeds the lane capacity. Still, the delay road users experience is not considered in the allocation of green time. Furthermore, non-recurrent spillback is not identified by the proposed method. In terms of the researched effects in the simulation, the simulation environment does not fully replicate real-life behaviour. Furthermore, the performed experiments assumed an exact and correct queue length retrieved from the simulation, and sensor faults or misestimation of queue length are not considered. Regarding the corridor, the network performance could be overestimated as no spillbacks can occur from directions that leave the corridor. The pedestrians are included in the simulation without adequate behaviour; additionally, the behaviour of the drivers and the environment itself is limited. Therefore, the effects may be underestimated or overestimated compared to a real-life setup.

This research led to several recommendations for future research. First, in terms of spillback detection, it is advised to research further the identification of non-recurrent congestion and the corresponding effects occurring from that congestion, such as the change in link capacity and handling that congestion in terms of traffic signal control. Second, regarding the priority provided to a spillback direction, the advice is to research further how spillback should be prioritized in different types of traffic situations in a multimodal context. Additionally, more insight can be gained in combining different priorities, such as the configuration of the vehicle priority in combination with other spillback priorities. Regarding spillback control strategies, it is advised to look into the current design and vary the weights given to the spillback penalty depending on the traffic situation. Regarding alternative control strategies, it is recommended to research other control strategies and their effect on the traffic and network performance, such as approaches with an alternative objective function when spillback is detected or including downstream and upstream metering in the solve method. In terms of policy, it is recommended to gain insight into the effects of different combinations of priorities (vehicle, movement, and spillback) to fulfil different policy goals, and the effects of these policy goals on the traffic and network performance. Altogether, when the policy goal is to reduce traffic fatalities to zero as soon as possible, it should be no surprise that controlling an emerging unsafe situation, such as spillover at an upstream intersection, should be highly prioritized in traffic signal control.

## Introduction

1

#### <span id="page-20-1"></span><span id="page-20-0"></span>**1.1. Research context**

Many configurations of traffic systems have been proposed, tested, and implemented to promote traffic flow and network performance in urban environments. When traffic demand is high and the intersection has a lack of throughput, phenomena that are paired with oversaturated conditions occur. For example, oversaturation of a turning-bay which leads to lane blockage. In addition, oversaturation of an intersection lane, which could lead to vehicles spilling over to an upstream intersection. These phenomena are called spillback effects and could lead to unsafe traffic situations. For example, when green time is allocated to a direction, while conflicting traffic is not yet sufficiently cleared. In the United States, some municipalities assign traffic safety a really high priority, such as the [San Francisco County Transportation](#page-84-0) [Authority](#page-84-0) [\(2018](#page-84-0)) has the vision of reducing traffic fatalities to zero by 2024. As in the United States, one-third of intersection fatalities still occur at signalized intersections([Stewart](#page-85-0), [2022](#page-85-0)), one way existing traffic systems could contribute to this goal is by ensuring unsafe situations that are paired with oversaturated conditions are avoided or handled faster by providing additional priority in the traffic light control.

#### <span id="page-20-2"></span>**1.1.1. Real-time traffic control systems and spillback**

Many real-time control traffic systems configurations have been around for quite some time. Such as as the Sydney Coordinated Adaptive Traffic System (SCATS) and the Split Cycle and Offset Optimization Technique (SCOOT) which were designed in the 1970s([Sims & Dobinson](#page-85-1), [1980](#page-85-1); [Bretherton,](#page-82-1) [1990\)](#page-82-1). Or relatively newer RTCS such as UTOPI-A/SPOT [\(Mauro & Di Taranto](#page-84-1), [1990;](#page-84-1) [Wahlstedt,](#page-85-2) [2013](#page-85-2)) or RHODES([Mirchandani & Head,](#page-84-2) [2001](#page-84-2)), which were designed in the early and late 1990s, respectively. Although varying in configuration and focus, RTCS have the same philosophy, namely switching traffic light signals or modifying a signal plan according to the optimization of a criterion for the next *n* seconds. The number of seconds depend on a configured time horizon, and every *m* seconds the horizon is renewed depending on the update frequency([Boillot et al.,](#page-82-2) [1992\)](#page-82-2). These RTCS have struggled to consider congested conditions due to uncertain traffic flows and cause wrong prediction, misestimation of the actual traffic state, or insufficient traffic control strategies [\(Boillot et al.,](#page-82-2) [1992](#page-82-2); [H. Liu et al.](#page-83-0), [2009;](#page-83-0) [Ramezani & Geroliminis,](#page-84-3) [2013](#page-84-3)). Over the years traffic control have been further developed, not only to control and optimize traffic in undersaturated conditions, but also for moderate and oversaturated traffic conditions([Wahlstedt,](#page-85-2) [2013\)](#page-85-2) (such as SCOOT and RHODES [\(Zargari et](#page-86-0) [al.](#page-86-0), [2016\)](#page-86-0)).

Traffic phenomena that occur with oversaturated conditions are spillback effects. When using loop detectors, spillback is mainly detected using lane capacity in various ways. Approaches use link capacity in terms of link length and queue length, the maximum allowed queue length, and maximum allowed occupancy. Another approach is detection of vehicle speed of a downstream vehicle or vehicle occupancy combined with a probability that estimates spillover probability. In terms of solving spillback, signal control strategies are used in two ways. The first strategy is a form of metering by limiting the inflow with a reduction in green time for critical phases at upstream intersections, the green time extension for the congested link, or both. The second approach is with an objective function. It uses an optimization criterion that minimizes a variable to solve spillback queues, such as minimizing the queue length, maximizing the throughput or minimizing the delay. The objective function optimizes the criterion so that the green time for the critical link is set to

green. On the other hand, traffic is not only prioritized for spillback control. With a traffic mix with multiple types of road users, public transit can also receive priority, and in urban areas also priority is provided for bicycles [\(Fietserbond](#page-83-1), [2022](#page-83-1)).

#### <span id="page-21-0"></span>**1.1.2. Multimodal real-time control systems**

In the development of RTCS, providing priority to transit was found to be effective. Transit Signal Priority (TSP) has been used to improve the level of service of transit users, with key performance indicators such as regularity and punctuality. TSP is incorporated into the signal control and modifies the normal signal operation process to accommodate transit vehicles([Baker et al.,](#page-82-3) [2002\)](#page-82-3). In combination with bus priority, SCOOT uses inductive loop detection and has proved effective and reliable([Hounsell & Shrestha](#page-83-2), [2005](#page-83-2)). In addition, with improved sensors (e.g. roadside beacons for the bus), TSP was demonstrated as more effective. RTCS with TSP for other modes, such as LRVs, have also been developed. Here the LRV has the additional priority and traveltime for the vehicle is minimized, while the delays for passenger cars do not increase [\(A. Stevanovic, Kergaye, & Martin,](#page-85-3) [2009](#page-85-3)). An improvement in sensor quality, from magnetic loop detectors to data-driven sensors such as video detectors or laser sensors, feed the exact vehicle location, speed, and trajectory to the traffic system. In combination with the improvement in computing power, data-driven models have become a valuable addition to RTCS. However, the impact of multimodal users in urban areas increases the complexity of finding optimal signal control because the RTCS and TSP systems often have conflicting control objectives [\(He et al.,](#page-83-3) [2014](#page-83-3)). Hence, RTCSs have been developed to optimize multiple road users into the control objective. The control objective is based on vehicle delay but could also be based on personal delay. The personal delay was found more successful in objective functions since it considers the passenger occupancy of a vehicle([Christofa](#page-82-4), [2012;](#page-82-4) [J. Stevanovic](#page-85-4), [2011](#page-85-4)). For multimodal RTCS that optimize vehicle delay, user-specified priority lists are included to account for difference in occupancy [\(Yagar & Han](#page-85-5), [1994;](#page-85-5) [J. Stevanovic](#page-85-4), [2011](#page-85-4)). For example, a light-rail vehicle (LRV) can be weighted 20 times more due to the number of passengers in the vehicle or the justification that the LRV has to adhere to the transit schedule.

#### <span id="page-21-1"></span>**1.2. Knowledge gaps**

Many real-time controlled systems have been proposed, compared, and assessed in the literature. Control strategies have been developed to detect, avoid or manage spillback effects. Previous studies on unimodal RTCS have improved control strategies to also control vehicles in more moderate and congested conditions, along an arterial or in a network, in which detection and control strategies for queue spillback have been found. With the rise of multimodal optimization in RTCS as the control objective, reduced delay and improved flow are perceived benefits([Van Katwijk](#page-85-6), [2008;](#page-85-6) [He et al.](#page-83-3), [2014](#page-83-3); [Mein et al.](#page-84-4), [2022](#page-84-4)). These systems have been tested and operate in undersaturated conditions. However, they have yet to be refined to work in moderate or oversaturated conditions. Therefore, looking into the feasibility of spillback control strategies in a multimodal context([Noaeen et al.,](#page-84-5) [2021\)](#page-84-5) can be valuable. It was recommended to incorporate detection methods and control strategies for spillback, as these multimodal systems currently have researched benefits in undersaturated traffic conditions([Yagar & Han,](#page-85-5) [1994](#page-85-5); [Christofa et al.](#page-82-5), [2013](#page-82-5); [He et al.,](#page-83-3) [2014;](#page-83-3) [Christofa et al.,](#page-82-6) [2016](#page-82-6); [Mein et al.](#page-84-4), [2022\)](#page-84-4). To research these components, research questions are proposed.

#### <span id="page-21-2"></span>**1.2.1. Scientific relevance**

There has been little research on spillback control strategies in real-time controlled systems with multimodal optimization. In unimodal RTCS [Noaeen et al.](#page-84-5) [\(2021](#page-84-5)) developed a control strategy that avoids and handles spillback effects, suggesting that more research is needed to include more modes and test transit priority in traffic control that incorporates spillback modelling. In the context of multimodal RTCS, it is suggested that a more sophisticated queue estimation model is used([He et al.,](#page-83-3) [2014\)](#page-83-3), which is done by [Christofa et al.](#page-82-6) [\(2016\)](#page-82-6) with a queue model that applies SWT. However, using SWT to estimate delays, not queues, leads to no detection of spillback and the possibility to control the identified spillback. Therefore, it is useful to gain insight into the effect of spillback control strategies on the traffic performance and network performance in a RTCS with multimodal optimization.

#### <span id="page-21-3"></span>**1.2.2. Societal relevance**

Research results lead to a more robust multimodal RTCS, which benefits the traffic participants in urban areas. First, in terms of throughput in which (transit) vehicles could achieve higher overall throughput leading to lower travel times for road users. Second, avoiding or handling unsafe situations that occur due to oversaturated traffic phenomena, as this leads to less conflicts between different road users. For example, when an intersection is not cleared when an approaching light-rail vehiclecrosses the street or when pedestrians cross the street when passenger cars are still halted at the crosswalk due to downstream congestion.

#### <span id="page-22-0"></span>**1.3. Research questions**

The research focus is to gain insight into the effects of spillback control strategies in a real-time controlled system with multimodal optimization on traffic and network performance. A spillback detection method and control strategy must be designed to gather insight, furthermore, the detection method and control strategy must be implemented in the RTCS. Thereafter, the designed approach should be tested on a simulation network. The knowledge gaps and the research focus result in the following research questions (RQs):

- 1. *How can an existing RTCS be extended with spillback detection and spillback control strategies in a multimodal context?*
	- 1.1 How could spillback be detected?
	- 1.2 How could spillback be avoided or handled?
	- 1.3 What is the architecture of a multimodal RTCS?
	- 1.4 How can 1.1 and 1.2 be designed in a RTCS with multimodal optimization?
- 2. *What is the effect of controlling spillback in a multimodal RTCS at signalized intersections on the traffic and network performance?*
	- 2.1 What is the effect of spillback control strategies on the traffic and intersection performance of an isolated intersection with a car flow, and a mixed traffic flow?
	- 2.2 What is the effect of spillback control strategies on the traffic and network performance in a corridor with a mixed traffic flow?

#### <span id="page-22-1"></span>**1.4. Overview of methodology**

To answer the research questions a literature review is conducted, an existing RTCS with multimodal optimization is extended with a spillback detection method and spillback control strategies, furthermore, a simulation study is performed. In [Figure 1.1](#page-23-2) an overview is given how these research steps have been performed and structured in the thesis. In the next subsections the research steps are described.

<span id="page-23-2"></span>

Figure 1.1: Overview research steps and methods

#### <span id="page-23-0"></span>**1.4.1. Literature review**

To fill in the knowledge gap, a literature review was conducted to gain insight into spillback control strategies in RTCS and also to identify the architecture of the used RTCS with multimodal optimization. First, relevant keywords were constructed for the literature review surrounding traffic control in RTCS which incorporate spillback control at signalized intersections. The keywords led to papers that research spillback detection methods or spillback control strategies in real-time controlled traffic systems, the collection was extended by forward and backwards snowballing. An overview of the paper trail is shown in [Table A.1](#page-87-1). Second, the found articles were evaluated using the PROMPT (Presentation, Relevance, Objectivity, Method, Provenance, Timeliness) approach. As the papers were found using a peer-reviewed database and found papers were part of a scientific journal or conference proceedings, most papers were according to standards regarding the presentation and provenance (e.g. origin) of the paper. In terms of relevance, it was inspected whether the papers proposed applicable spillback detection methods and control strategies. This led to a first collection of relevant papers. The resulting collection was further analyzed in terms of objectivity, but no papers from the collection were dropped from this.

The literature review led to a collection of papers with approaches to detect and control spillback control strategies. This collection was used as input for the design of spillback component, in which there was evaluated whether the method was suitable as the extension of an existing RTCS with multimodal optimization.

#### <span id="page-23-1"></span>**1.4.2. Design of spillback component**

The literature review generated a collection of approaches to detect and control spillback (RQ1.1 and RQ1.2). The identified architecture of the multimodal RTCS (RQ1.3) provided a framework to implement the methods in. The design of the spillback components consisted of two parts. First, the component design as an extension to the RTCS, second, the implementation of the spillback component to a test environment.

For the test environment, the traffic system was translated into a software-in-the-loop simulation (SILS). As the system

controls traffic light, based on individual vehicle delay, the test framework uses a microsimulation. The test framework of MobiMaestro-Flow already existed in this framework and were extended by the spillback detection method and spillback control strategies, in which the spillback detection queue lengths were retrieved directly from the simulation, assuming perfect information about the queue length. In this environment experiments were performed and data was generated by the simulation environment SUMO.

#### <span id="page-24-0"></span>**1.4.3. Simulation**

#### **Data collection**

Data for the analysis was generated by the simulation-software. More specific data, such as the vehicle speed and locations were generated by detectors placed in the simulation network. For example, to measure the duration of spillback, a detector was placed upstream of a turning bay to detect lane blockage.

Data from two simulation networks was used. The first network was an isolated intersection; the second simulation network was a corridor consisting of ten intersections. The collected data was filtered and processed to retrieve relevant variables. These relevant variables were used in the assessment to determine the effects of spillback control strategies. The key performance indicators (KPIs) for traffic performance were chosen as the delay, and the spillback duration. The spillback duration was chosen because it shows a direct effect of the spillback control strategy; furthermore, in directions where no spillback could occur, the duration resembles the queue duration. The delay is included to assess the impacts on the vehicle routes when a conflicting, or non-conflicting direction, receives additional spillback priority and therefore more green time. The delays is derived from travel time. For the network performance, it was chosen to include the network delay which is the sum of the individual intersection delays.

#### **Analysis**

For each scenario, the data consisted of the means and the standard deviation of every replication from the scenario. For the analysis ten replications of each scenario were compared. The traffic light control consisted of an adaptive traffic signal control that minimized the waiting time for all road users and was constrained by the ring-and-barrier structure. Experiments were carried out without the spillback control strategy (base case), and also with the incorporation of the spillback control strategy. The KPI means of these scenarios were compared. To test statistical significance, a Welch t-test was performed to test whether the means are significantly different with a two-tailed significance.

<span id="page-24-1"></span>The methodology covers multiple chapters. The structure of the thesis is depicted in [Figure 1.1.](#page-23-2)

#### **1.5. Thesis outline**

The thesis consists of eight chapters. After the introduction, the literature is reviewed in [chapter 2,](#page-25-0) starting with stateof-the-art RTCS, followed by spillback detection methods, spillback control strategies, transit signal priority (TSP), and RTCS with multimodal optimization. After that, the utilized multimodal RTCS is described in [chapter 4.](#page-46-0) The output of the literature review is used as input for the design of the spillback component, which is described in [chapter 4.](#page-46-0) Furthermore, the implementation of the spillback component in the test framework is described in [chapter 5.](#page-51-0) In this test framework multiple experiments are performed which follow the experimental design described in [chapter 6.](#page-58-0) The results of the experiments are presented in [chapter 7,](#page-67-0) followed by the discussion of the results. In [chapter 8](#page-77-0), the research questions are concluded and the limitations of the research are described. Lastly, the recommendations are done in terms of future research topics, policy, and model extensions.

# $\bigg)$

## Literature review

<span id="page-25-0"></span>This chapter describes the results of the conducted literature review. The main goal of the literature review is to gain insight into spillback detection methods, and spillback control strategies as an extension to multimodal RTCS. First the research approach is described. After that, state-of-the-art RTCS are described after which spillback detections methods in RTCS are reviewed. As RTCS with multimodal optimization are part of systems that provide signal priority, an outline is given of the development in (transit) signal priority. Next, insight is provided on spillback control strategies in comparable real-time controlled traffic systems. The following sections describe the developments regarding transit signal priority in these systems, which lead to the sections surrounding real-time control with multimodal optimization, and the lack of spillback control in these systems. The found results lead to a conceptual framework depicted in[Table 2.2](#page-38-1) at the end of this chapter. Finally, the findings of the literature review are concluded.

#### **Approach**

First, [Mein et al.](#page-84-4) [\(2022](#page-84-4)) and the references were analyzed to gain knowledge in the literature on RTCS with multimodal optimization and transit signal priority and identify the architecture of a multimodal RTCS. Subsequently, relevant keywords were constructed for the literature review about traffic control in RTCS, incorporating spillback control at signalized intersections. Keywords used for the literature review for spillback detection and control strategies are depicted in [Table 2.1](#page-25-1).

<span id="page-25-1"></span>Table 2.1: Generated search queries for literature study



#### **Search strategy**

The queries were entered in Scopus to find peer-reviewed literature surrounding spillback control strategies. First, ad-

ditional keywords were added to the query to narrow search results. This led to too few papers. Query 5, with 54 hits, was used as a starting point to review the literature.

The found articles were scanned and filtered, as explained in [subsection 1.4.1.](#page-23-0) The results consisted spillback detection methods and control strategies, and insight into the architecture of a multimodal RTCS. These papers were analyzed and distinguished according to the following characteristics: included road users in the proposed model, vehicle priority, objective function used for traffic light signal control (if any), and used traffic sensors for traffic detection. Furthermore, the utilized queuing model (if any), the type of spillback detection, and the spillback control strategy. The consideration between the different methods aims to answer research questions 1.1 to 1.3. An overview of the paper trail can be found in [Table A.1](#page-87-1) in [Appendix A,](#page-87-0) furthermore, the literature review led to a conceptual framework, which is fully depicted in [Table A.2](#page-88-0). In the next sections the literature is reviewed.

#### <span id="page-26-0"></span>**2.1. State-of-the-art RTCS**

Real-time traffic control systems can be divided into two categories. The first category is making minor changes to a previously defined signal plan; the second is more dynamic by switching traffic lights to red or green at each timestep. Despite this categorization, the configuration and focus vary, such as minimum green times or red time clearance. A part of this state-of-the-art real-time controlled traffic systems share the same philosophy: switching traffic light signals or modifying a signal plan according to an optimization criterion for the next*n*seconds. The number of seconds depends on a configured time horizon, and the horizon renews every *m* seconds according to a specified update frequency([Boillot](#page-82-2) [et al.,](#page-82-2) [1992](#page-82-2)). For example, an optimization criterion is usually the total delay at the intersection of multiple intersections in the network during the time horizon, which is approached by the sum of the individual car delay lengths over each timestep on the controlled links. Additionally, other optimization criteria have been developed, such as the sum of vehicle stops, delay experienced per person (personal), maximum vehicle flow, or combinations of other indicators.

<span id="page-26-1"></span>

Figure 2.1: Conceptual framework of Real-time control traffic systems

A RTCS of the first category is SCOOT, a traffic signal control system developed in the 1970s that modifies existing traffic signal settings based on the sum of the stops and the saturation degree([Bretherton](#page-82-1), [1990;](#page-82-1) [Robertson & Brether](#page-84-6)[ton,](#page-84-6) [1991\)](#page-84-6). Like SCOOT, SCATS considers adjustments of various signal timing separately; however, SCATS adjusts signal timings based on changes in traffic flows from previous cycles. Collected sensor data measures the Degrees of Saturation (DS) and Link Flows (LF) to allocate green time per direction([Sims & Dobinson](#page-85-1), [1980\)](#page-85-1). The system then calculates the cycle length, phase splits, and offsets([Boillot et al.,](#page-82-2) [1992;](#page-82-2) [A. Stevanovic, Kergaye, & Martin](#page-85-3), [2009\)](#page-85-3). Some examples in the second category of RTCS, deciding whether to switch the traffic light or not, are The Optimize Policies for Adaptive Control (OPAC), developed by [Gartner et al.](#page-83-4) ([2001](#page-83-4)), and CRONOS developed by [Boillot et al.](#page-82-2) [\(1992\)](#page-82-2). Updates of OPAC realized enhancements with features such as full intersection simulation with individual and platoon vehicle identification, performance functions for total delay or stops of intersections, cycle length and offset optimization, mode coordination, and dynamic phasing. Furthermore, CRONOS controls zones of multiple intersections. In addition, it utilizes video sensors instead of conventional loop detectors for traffic flow measurements and handling oversaturation traffic phenomena [\(Boillot et al.,](#page-82-2) [1992](#page-82-2), [2006\)](#page-82-7). The forecasting module in CRONOS predicts the future arrivals on each link, which uses the simulation module and calculates the optimization criterion for traffic signal states

(e.g. a red or green light status). The following time step identifies traffic signal states to choose an optimal sequence. Furthermore, large-scale system control systems were developed to control multiple traffic participants, such as UTOPI-A/SPOT [\(Mauro & Di Taranto,](#page-84-1) [1990\)](#page-84-1). The main features of UTOPIA/SPOT are priority assignment to public transit and optimization of private vehicles in all traffic conditions. UTOPIA/SPOT uses local minimization of a cost function for each intersection using a SPOT unit to optimize signal timings. The SPOT units from the systems communicate the counted and predicted traffic flows and signal changes. After that, each controller determines the cycle coordination for each intersection. All in all, for each RTCS, traffic sensors need to supply the RTCS with sufficient information for optimal signal control.

#### <span id="page-27-0"></span>**2.1.1. Traffic sensors**

RTCS collect infrastructure data and traffic object data, such as the status of a traffic light, or the location or speed of a vehicle. The type and amount of data collected for the RTCS depend on the used traffic sensors. For example, magnetic loop detectors as sensors upstream of an intersection can count vehicles, approximate vehicle speed, and the time gap between cars. However, more precise data, such as the vehicle gap and speed, are more sensitive to sensor faults, such as a wrongly parked car on the side of the road. Nevertheless, RHODES uses these upstream detectors to predict arrivals and includes these future arrivals in the queue length prediction to allow for a longer prediction horizon([Mirchandani &](#page-84-2) [Head](#page-84-2), [2001](#page-84-2)). But also, OPAC uses upstream-placed detectors to obtain arrival data for the head portion of the prediction [\(Gartner et al.](#page-83-4), [2001\)](#page-83-4).

The loop detectors used by RTCS are interchangeable with traffic sensors, enabling RTCS to receive more accurate data. For example, traffic objects can be identified using LiDAR sensors([J. Wu et al.](#page-85-7), [2020\)](#page-85-7) or cameras([Albiol et al.](#page-82-8), [2011](#page-82-8)), or travel times measured by cameras can be used for arrival predictions in queue length estimation([Ma et al.](#page-84-7), [2018](#page-84-7)). Furthermore, CRONOS uses machine vision to incorporate spatial information into estimating traffic densities, flows,and speed, such as the number of vehicles in the inner junction of an intersection ([Boillot et al.,](#page-82-7) [2006\)](#page-82-7). Datadriven traffic models use data to promote a better traffic flow. However, when it is raining or misty, the reliability of these sensors reduces drastically; bad weather will also reduce the sensor range for a precise approximation. However, recent research showed a deep learning approach where a queue length is estimated using convolutional neural networks [\(Umair et al.,](#page-85-8) [2021](#page-85-8)), making the estimation of traffic variables possible with a limited-quality video input.

Ultimately, data from the traffic sensors is collected and processed so that the data can be used for modelling, namely the estimation and prediction of traffic variables, such as the traffic flow or the vehicle density. .

#### <span id="page-27-1"></span>**2.1.2. Modelling**

In the modelling part of a RTCS, specific variables are predicted or estimated, dependent on the configuration of the RTCS. For example, SCOOT calculates the occupancy and vehicle flows for the allocation of green time, SCATS uses the traffic flows from previous cycles, and OPAC uses performance function for total delay or vehicle stops of intersections. Furthermore, CRONOS also uses the delay as an optimization criterion, in which queues are stored as the number of vehicle on the link.

The queue, or the number of vehicles on the link, used for the delay calculation are often calculated based on an inputoutput method, as shown in [Equation 2.1.](#page-27-2) The number of vehicles in the next time step *t* + 1 is equal to the number of vehicles in the queue during time step *t* plus vehicles flowing in, the arrivals *A*(*t*), minus vehicles flowing out, the departures *D*(*t*). The vehicles in the queue can then be described using the number of vehicles, but can also be described in meters, using the vehicle length and the gap between two consecutive vehicles.

<span id="page-27-2"></span>Vehicles in queue
$$
(t + 1)
$$
 = Vehicles in queue $(t) + A(t + 1) - D(t + 1)$  (2.1)

In the next subsections, traffic variables that are commonly estimated or predicted, such as the traffic flow, the delay, and the queue length.

#### Traffic flow and density

For the estimation of arrivals, the expected arrival time is predicted using upstream detectors, the time-instant is recorded, and the arrival time is estimated using the travel time of the vehicle*i*, which travels over the sensor during time step *t*, to the downstream intersection, as shown in [Equation 2.2](#page-27-3).

<span id="page-27-3"></span>Arrival time(*i*, *t*) = Arrival time detector(*i*, *t*) + distance of upstream-detection/
$$
v_{freeflow}
$$
 (2.2)

The number of departures during a cycle depends on the allocated green time  $T_g^n$  to the vehicles in queue. The discharge of vehicles is dependent on the queue discharge rate of the vehicles. The queue discharge rate is dependent on the utilized queue method. For example, in [Equation 2.3](#page-28-1) the number of departures are calculated using a queue discharge rate that is equal to the saturation headway of the vehicles, furthermore, a start-up delay is included for the first vehicle  $(i = 1)$ . The upper limit *I* is the number of vehicles that are able to depart during the allocated green time.

<span id="page-28-1"></span>
$$
\text{Departures}(i, n) = (\text{start-up delay} + \sum_{i=2}^{I} \text{ vehicle}(i) \cdot \text{Queue discharge rate}) \le T_g^n \tag{2.3}
$$

The expected arrivals and departures can also be used for the density calculation. For example, SCOOT uses the density and gives the factor a degree to determine the weight of the density compared to the available road capacity (Degrees of Saturation). In combination with the vehicle flows, this is used in the green time allocation.

#### Queues

<span id="page-28-0"></span>The estimated number of departures, but also the vehicle delay computation is dependent on the utilized queue method. Two types of queuing models are widely used. On the one hand, queuing models that neglect the spatial characteristics of a queue (vertical queue). On the second hand, models that take spatial variables into account to estimate the head and tail of the queue (horizontal queue), as depicted in [Figure 2.2](#page-28-0).



(spatial characteristics taken into account)

Figure 2.2: Horizontal Queue vs. Vertical queue

A vertical queuing model considers a standing queue where vehicles are stacked virtually, the cars are stacked at the bottleneck's place, in the case of an intersection, at the stopbar. Consequently, the queue length has no influence on approaching vehicles [\(Y. Liu et al.,](#page-83-5) [2018](#page-83-5)). The queuing process consists of a deterministic queue profile: at the queue's release, the queue's bottom is released first, following a First in First Out method (FIFO). For the queue discharge, a queue discharge rate is set based on the vehicle headway, the acceleration rate, or a combination of both. For the first vehicle a (fixed) start-up delay is taken into account for the driver's responsiveness.

A HQM describes a queue with, for example, the cell transmission model, in which the road is divided in cells to describe kinematic wave equations([Daganzo,](#page-82-9) [1995\)](#page-82-9). Another widely used approach is applying a shockwave queuing profile developed by [Lighthill & Whitham](#page-83-6) [\(1955\)](#page-83-6), which estimates the head and tail of the queues by describing time and spacedependent dynamics. When SWT is used as a queue method, a queue profile is constructed using the traffic flow, density, and speed. These three variables are referred to as traffic states, in which vehicles can transition between uncongested and congested states, according to a corresponding density, speed, and flow. In a fundamental diagram, points are used to describe traffic flow phenomena, such as the queue built up and the queue discharge process. The fundamental diagram (see [Figure 2.3a](#page-29-1)) and shockwave profile (see [Figure 2.3b](#page-29-1)), show the queue process, according to allocated green time, red time, the queue built up, and queue discharge. The numbers indicated in [Figure 2.3a](#page-29-1)are corresponding to the indicated numbers in [Figure 2.3b,](#page-29-1) and resemble a traffic state – for example, state 2, indicates a jam density with a vehicle flow of zero.

<span id="page-29-1"></span>

Figure 2.3: A fundamental diagram and a queue profile, according to SWT

The VQM and HQM are different in operational mechanisms but are both vehicle-trajectory oriented and operable using fixed-location sensors. However, both have a lack of operability when sensors are faulty or insufficient information is available([Y. Liu et al.,](#page-83-5) [2018\)](#page-83-5). In that case, the benefit of choosing a VQM above an HQM is computational efficiency. On the other hand, an HQM has the benefit of describing spatial queue characteristics, with advantages such as implementing the queue length in the objective function, a check whether a queue length exceeds a lane threshold, and delays of long queues are less underestimated compared to VQMs.

#### Delay calculation

The delays computation depends on the queue method of RTCS. For delays, using a VQM, are often comparable, such as the approaches from [Sharma et al.](#page-84-8) [\(2007](#page-84-8)); [Boillot et al.](#page-82-7) [\(2006](#page-82-7)); [He et al.](#page-83-3) ([2014](#page-83-3)); [A. Stevanovic et al.](#page-85-9) ([2015\)](#page-85-9); [Bhouri et](#page-82-10) [al.](#page-82-10) [\(2015](#page-82-10)), with a fixed queue discharge rate, based on the start-up delay and configured vehicle headways. This leads to the following equation, in which *I* is the total number of vehicles in the queue:

Delay(*i*) = start-up delay + 
$$
\sum_{i=2}^{I}
$$
 vehicle(*i*) · Queue discharge rate (2.4)

Another approach, which takes spatial characteristics into account, is by applying shock wave theory. The shockwave profiles, as in [Figure 2.3b](#page-29-1), can be used to estimate collective delays experienced by vehicles. The delay is calculated as the time vehicles wait for a red traffic light. Assuming that the cars arrive in platoons without dispersion, [Christofa](#page-82-4) [\(2012\)](#page-82-4) uses the arrival time of the first vehicle, the platoon size, and the residual queue of the upstream intersection of the previous traffic light cycle to estimate the collective delay, an example is depicted in [Figure 2.4.](#page-30-2)

According to [Figure 2.4,](#page-30-2) the individual vehicle delay  $d$  of vehicle  $i$  during cycle  $c$  can be estimated using the time  $t_i^c$  the vehicle *i* arrives at the intersection, during cycle *c*, and the time  $t_i^c$  the vehicle *i* is expected to leave the intersection at  $\csc c + 1$ :  $t_i^{c+1}$ .

$$
d(t)_i^c = t_i^{c+1} - t_i^c \tag{2.5}
$$

The individual vehicle delays are then summed to calculate the collective vehicle delay during cycle *c*:

$$
D(t)_i^c = \sum_{i=1}^I d(t)_i^c
$$
 (2.6)

<span id="page-29-0"></span>In the modelling part of a RTCS, the estimated or predicted traffic flows, densities, queue length, or delays are calculated so that they can be utilized for the optimization of an objective function.

<span id="page-30-2"></span>

Figure 2.4: Delay computation in a shockwave profile

#### **2.1.3. Optimization**

The modelled traffic variables, such as the vehicle delay, is used in the optimization of the RTCS. The optimization is also dependent on the chosen signal control, in which existing traffic timings are adjusted with minor changes (e.g. SCOOT), or a signal plan can be adjusted dynamically by switching to red or green every time step (e.g. OPAC, CRONOS), with an adaptive approach that also considers phase skipping. The optimization consists of the minimization or the optimization of the objective criterion, which is found by different search algorithms.

In some cases an approach that searches brute force through the possible signal plans is used. For example, when the traffic network is small, or the signal plans are already predefined. But also search algorithms for local optima are used, to increase the update frequency or to expand the optimization horizon. A disadvantage is that the found signal plan may not the optimal signal plan, but the optimization time for a new traffic signal plan is minimized. For example, CRONOS, finds new traffic signal states for the next second, in less than a second([Boillot et al.,](#page-82-7) [2006](#page-82-7)).

#### <span id="page-30-0"></span>**2.1.4. Spillback effects**

When demand is high, traffic phenomena occur that come with oversaturated conditions. For example, when a queue grows, it spills over to the upstream intersection. Spillover could lead to blocking other passenger cars at the intersection and other travel modes, such as light-rail vehicles. Another effect is lane-blockage, for example, when an oversaturated turning bay blocks straight-through traffic. Furthermore, the straight-through lane could also be oversaturated, causing lane blockage to the turning-bay [\(Boillot et al.](#page-82-2), [1992;](#page-82-2) [H. Liu et al.](#page-83-0), [2009;](#page-83-0) [Ramezani & Geroliminis,](#page-84-3) [2013](#page-84-3)). With that in mind, more situations could occur due to spillback, such as blocking pedestrians when vehicles spill over to the upstream intersection. Examples of these spillback effects are depicted in [Figure 2.5](#page-31-0)

Unsafe situations occur with spillback effects, such as when a passenger car spills over to an upstream intersection and blocks an incoming Light-Rail vehicle. Additionally, when cars are blocking other road users – for example when cars cannot travel further downstream as other vehicles spill back to the intersection. Consequently, these cars block other more vulnerable road users such as cyclists and pedestrians.

<span id="page-30-1"></span>In RTCS traffic variables are taken into account that measure performance criterions of cars, but when RTCS can identify spillback situations, the system could provide additional priority to solve situations as in [Figure 2.6,](#page-31-1) so that straight traffic have improved throughput. Therefore, unsafe situations that can occur with spillback, such as conflicts between different types of road users, are avoided. The following section describes spillback detection and solving methods of RTCS.

<span id="page-31-0"></span>

Figure 2.5: Examples of spillback effects at a signalized intersection

<span id="page-31-1"></span>

Figure 2.6: Spillback effects and conflicts with multiple road users on the Burgermeester Jamessingel

#### **2.2. Spillback control in RTCS**

Several approaches have been researched and developed to detect and handle spillback effects in RTCS. In these papers, several characteristics that vary in the detection and handling method of spillback effects have been identified. In terms of spillback detection and control strategies, the components included road users, vehicle priority, the utilized objective function (if any), the type of traffic signal control, used traffic sensors, and queue method (if any). In the next sections, these components are described, according to the found literature.

#### <span id="page-32-0"></span>**2.2.1. Spillback detection**

The spillback detection method is dependent on the utilized traffic sensors and if a queue method is used. In literature, two types of spillback are defined in detection: lane blockage and vehicles that spill over to an upstream intersection. The lane blockage can be a straight-forward direction blocking a turning-bay, or the other way around. Spill over is often used in terms of overflowing to an upstream intersection.

The following approaches were used often. First off, spillback is detected using a link capacity by defining a capacity in terms of the number of vehicles that a turning bay could handle without blocking another direction or the capacity of a link without spilling over to an upstream intersection [\(Y. Zhang & Tong](#page-86-1), [2008](#page-86-1); [Y. Liu & Chang](#page-83-7), [2011;](#page-83-7) [Han & Gayah,](#page-83-8) [2015;](#page-83-8) [Ramezani et al.](#page-84-9), [2016\)](#page-84-9). The traffic system obtains the number of vehicles by counting the departures at an upstream intersection or the passing vehicle at an advance detector. The second mainly used approach was comparable: a queue model estimates the queue length for the traffic light cycle. When the queue length exceeds the queue threshold, spillback is detected [\(Christofa et al.,](#page-82-5) [2013;](#page-82-5) [Wong & Lee](#page-85-10), [2020](#page-85-10); [Mohajerpoor & Cai,](#page-84-10) [2020;](#page-84-10) [Noaeen et al.,](#page-84-5) [2021;](#page-84-5) [H. Zhang](#page-86-2) [et al.](#page-86-2), [2020\)](#page-86-2). The queue threshold can be a turning-bay length or the link length of two consecutive intersections. Additionally, there are three other methods found. [X. Wu et al.](#page-85-11) [\(2010\)](#page-85-11) proposed an additional oversaturation index: the time a vehicle spends in a queue and the length of the queue determines whether spillback is sustained and should be dealt with. Or whether the spillback is acceptable as the vehicle moves further to the downstream intersection. Another approach that uses an upstream detector, is the method by [Ren et al.](#page-84-11) [\(2017](#page-84-11)), spillback is detected by finding a low speed at this upstream detector, with effective result, however it is very sensitive to sensor faults. Another approach that does not require estimation, is when problem areas, such as the inner junction of an intersection, or turning bays, are tracked using data-driven traffic sensors, such as cameras or laser sensors([Boillot et al.](#page-82-7), [2006\)](#page-82-7).

#### Traffic sensors

The used traffic sensors have an impact on the approach used for queue estimation and spillback detection methods. The more information an RTCS receives about the vehicle state, the more accurate the detection and handling of spillback is applied. For example, the systems from [Y. Zhang & Tong](#page-86-1) ([2008](#page-86-1)); [Y. Liu & Chang](#page-83-7) ([2011\)](#page-83-7); [Ramezani et al.](#page-84-9) [\(2016\)](#page-84-9); [Wong & Lee](#page-85-10) ([2020\)](#page-85-10); [Mohajerpoor & Cai](#page-84-10) [\(2020](#page-84-10)) assume perfect information about the queue length, vehicle position, and lane-changing behaviour. However, outcomes could overestimate the effects measured by the spillback control and handling strategies as perfect information is available. Other developed systems use connected vehicles in a connected vehicle environment([Christofa et al.,](#page-82-5) [2013;](#page-82-5) [Ramezani & Geroliminis](#page-84-12), [2015](#page-84-12); [Cao et al.,](#page-82-11) [2019;](#page-82-11) [H. Zhang et al.,](#page-86-2) [2020\)](#page-86-2), with advantages such as a more accurate and efficient traffic control, but also assumes accurate vehicle position, speed, turning behaviour, and vehicle gaps. As the approaches need connected vehicles, the approaches are less likely to be applied in the short term. Designs that have a calibrated traffic sensor, such as upstream loop detectors or stopbars, have the advantage that it is applicable in the short term in an urban environment. However, the approaches inliterature that only use detectors, also require a queue method to detect potential spillback.

#### Queue model

Queuing models were used to detect the maximum (and residual) queue length in real time to detect whether a queue threshold was exceeded. The most widely used queue model for spillback detection is the method that applies shock wave theory to estimate the maximum queue length [\(X. Wu et al.](#page-85-11), [2010](#page-85-11); [Christofa et al.](#page-82-5), [2013;](#page-82-5) [Han & Gayah](#page-83-8), [2015](#page-83-8); [Ramezani](#page-84-12) [& Geroliminis,](#page-84-12) [2015;](#page-84-12) [Noaeen et al.](#page-84-5), [2021](#page-84-5); [Mohajerpoor & Cai,](#page-84-10) [2020;](#page-84-10) [H. Zhang et al.](#page-86-2), [2020](#page-86-2); [Ren et al.,](#page-84-11) [2017\)](#page-84-11). Other approaches measured the queue length directly from the simulation, such as [Y. Zhang & Tong](#page-86-1) [\(2008](#page-86-1)); [Y. Liu & Chang](#page-83-7) ([2011\)](#page-83-7); [Chen & Chang](#page-82-12) [\(2014](#page-82-12)); [Ramezani et al.](#page-84-9) [\(2016](#page-84-9)). But also traffic variables from connected vehicles were used, such as the approaches from [Christofa et al.](#page-82-5) [\(2013\)](#page-82-5); [Ramezani & Geroliminis](#page-84-12) ([2015](#page-84-12)); [Cao et al.](#page-82-11) [\(2019](#page-82-11)). Another method by [Wong & Lee](#page-85-10) ([2020\)](#page-85-10), was a cell transmission model (CTM) around the intersection to describe queue dynamics. The CTM made it possible to optimize traffic signals, including the red and green time duration, cycle time based, in which lane markings on the road are included as binary optimization variables. Also, a gap-based queuing model from [Christofa](#page-82-5) [et al.](#page-82-5) [\(2013](#page-82-5)) was researched, but proved to be less accurate than applying shockwave theory.

#### Traffic signal control

In the articles, regarding spillback control with RTCS mainly two types of traffic signal control were used. The first is fixed time control, in which the traffic lights follow a predefined signal plan. The other type is adaptive control, which controls the traffic lights according to an optimizing an objective criterion. For fixed time traffic signal, the spillback control strategy interferes with the signal plan, and for adaptive control, the spillback control strategy can alter the objective criterion – for example, the objective function is changed to minimizing the total time spent on an intersection. Or the control strategy interferes with the optimal signal control. These strategies are explained in the next section.

#### <span id="page-33-0"></span>**2.2.2. Spillback control strategies**

In the literature, two types of spillback control strategies have been found. The first one solves spillback by an objective criterion. The other approach is metering upstream, downstream, or both at the congested intersection.

#### Objective function

In some of the researched articles, systems used an objective function to control the traffic. Approaches that utilize an objective function that detect spillback, uses additional delay, reduced link capacity for the congested intersection, and impacts on upstream intersections as new variables in the optimization. [Y. Liu & Chang](#page-83-7) [\(2011](#page-83-7)) use the objective function that maximizes the flow using a genetic algorithm and checks every step if oversaturation is present until the stop criterion is met. When oversaturation is found, a fitness evaluation determines whether the solution is near the optimal solution. The research from [Chen & Chang](#page-82-12) ([2014](#page-82-12)) builds further and incorporates heavy mixed traffic flows in the optimization. [Wong & Lee\(2020](#page-85-10)) also maximizes the flow for every direction in which the optimization prevents overflow. Another approach is that less capacity is assigned in the optimization to the link affected by spillback([Han &](#page-83-8) [Gayah,](#page-83-8) [2015\)](#page-83-8). The last type of optimization calculates the saturated green time (when a queue is fully discharged during a cycle) for all movements [\(Noaeen et al.](#page-84-5), [2021](#page-84-5)). The optimization minimizes the saturated green time for each possible stage and selects the phase with the maximum outflow rate for the phase time. After that, stages are improved by finding the set of phases with more movements compared to the selected stage. After that, the phase with the maximum outflow rate and the total queue length among the selected phases is chosen. Lastly, extended green is facilitated to allow flow continuation if a movement has received green in the previous and current phases.

#### Metering

Metering is restricting or promoting a flow in a specific direction to reduce or promote the incoming or outgoing traffic. In terms of metering, metering upstream is used to solve spillback by limiting the inflow to the congested link([Christofa](#page-82-5) [et al.,](#page-82-5) [2013;](#page-82-5) [Mohajerpoor & Cai,](#page-84-10) [2020\)](#page-84-10). Upstream metering is combined with downstream metering to promote downstream intersection flow and restrict upstream intersection flow [\(Ramezani et al.](#page-84-9), [2016](#page-84-9); [Cao et al.](#page-82-11), [2019\)](#page-82-11). Additionally, sidestreets can be added in upstream metering to limit the inflow even more([Ren et al.,](#page-84-11) [2017](#page-84-11)).

Multiple control strategies are developed to help cope with spillback effects and handle congested conditions. In these papers, the research focus has mainly been on the urban environment where only passenger cars are present. However, in an urban environment, there are mixed traffic flows, such as LRVs and pedestrians. State-of-the-art RTCS have been tested with vehicles and the addition of priority for public transit, such as SCOOT with bus priority([Hounsell &](#page-83-2) [Shrestha,](#page-83-2) [2005](#page-83-2)). Transit signal priority can improve the overall network performance of an urban environment. In the next section, Transit signal priority will be explained further.

#### <span id="page-33-1"></span>**2.3. Transit Signal priority**

There are many benefits to using transit signal priority control in traffic control systems. These benefits include improved transit schedule reliability, reduced transit travel times, improved level of service, and reduced stops for all modes, leading to increased travel comfort for both cars and transit users. Furthermore, reduction in emissions and, above all, increases the attractiveness of public transit, which is the consequence of priority competitiveness between public transit and cars.

In general, a physical TSP system comprises three main components – first, the vehicle detection system, which detects transit vehicles and generates priority requests. Secondly, receiving and processing the priority request at signalized intersections with the traffic signal control system. Lastly, the communication system connects the detection system with the traffic signal control system([Ding et al.,](#page-82-13) [2013](#page-82-13)).

Strategies for controlling TSP depend on the type of priority given to the traffic modes. Some strategies modify normal signal operations to allow transit vehicles to travel through a signalized intersection with delay; other systems are optimized, taking multiple modes into account for signal optimization. In the upcoming subsections, the fundamentals are explained based on the clear overview of [Baker et al.](#page-82-3) [\(2002](#page-82-3)). First, the different priority types are explained, as the difference between local TSP and area-wide TSP, the detection of transit in TSP, and lastly, the links between real-time control and TSP.

#### <span id="page-34-0"></span>**2.3.1. Priority types**

First, the difference between priority and preemption should be clear. The terms priority and preemption are different processes but are often confused: signal priority *modifies* the normal signal operation process to accommodate transit vehicles better, and preemption *interrupts* the standard procedure for special events([Baker et al.,](#page-82-3) [2002](#page-82-3)). In the case of urban intersections, these special events typically are the arrivals of emergency vehicles.

In TSP, multiple strategies exist to modify signal operations to allow transit vehicles to travel through a signalized intersection with a reduced delay. These can be distinguished mainly into three types: passive priority, active priority, and real-time priority, where the latter is the case when TSP is facilitated with improved information by enabling real-time detection information to be used for purposes other than real-time control. A schematic overview of priority types can be found in [Figure 2.7.](#page-35-3)

#### Passive priority

When transit operations are predictable and reliable, characterized by consistent dwell times and a high schedule adherence, in combination with high frequencies and low traffic volumes, passive priority strategies can be an efficient form of TSP. The signals coordinate transit vehicles' flow, promoting straight-through public transit. However, the consequence of traffic signals being coordinated for transit and not for other traffic modes is that these other may experience unnecessary stops or delays, which could lead to annoyance by other road users.

#### Active priority

Active priority is possible when public transit is detected in real-time. According to this detection, modifications are made to the traffic signal plan. These modifications are *early green*, which can be applied when the signal is red for the approaching vehicle. The green times for previous phases are often reduced to early green. The second strategy is *green extension*, where the downstream traffic light is green, and the method logically extends the green time to benefit transit priority. Optionally, early green and green extensions are combined for additional priority. *Actuated transit phases* detects transit vehicles at intersections. An actuated transit phase is a dedicated phase for transit vehicles. An actuated transit phase could be realized in two ways, either by *phase insertion*, in which the actuated transit phase is inserted in the normal signal sequence. The other way is to use *phase rotation* to provision TSP. An example of phase rotation would be a northbound left-turning bus requesting priority before the start of the green phase of the opposing through movement, which was the 'normal' phase structure [\(Baker et al.,](#page-82-3) [2002\)](#page-82-3).

#### Real-time control

When TSP strategies provide priority while simultaneously optimizing given performance criteria, such as person delay, transit delay, vehicle delay, or a combination dependent on the configuration of the real-time control system, generally, these systems require travel information about public transit for early detection to provide transit priority while minimizing traffic impacts for other modes. It is important to note that TSP does not end when the vehicle has passed through the signal. In priority strategies where the traffic system must detect the transit beforehand, most signal controllers implement a recovery operation where phases transition back to regular signal operation as stages were cut short or skipped during the priority strategy([Baker et al.,](#page-82-3) [2002\)](#page-82-3). This is the case because transit provision could negatively impact traffic operations of other modes.

<span id="page-35-3"></span>

Figure 2.7: Overview of priority strategies (adapted from [\(Yagar & Han](#page-85-5), [1994\)](#page-85-5))

#### <span id="page-35-0"></span>**2.3.2. Local TSP vs Network TSP**

By detecting approaching transit vehicle upstream of signalized intersection TSP could be accomplished at local intersection levels. The upcoming bus or LRV sends a call or message to the traffic signal controller to check in. After the vehicle has passed through the intersection, it can be detected again when it sends a call or message to check out.

At the network level, a traffic system uses an automated vehicle location and control (AVLC) system to determine if the transit vehicle travels on schedule. When the public transit is not travelling on schedule, the traffic signal controller could respond to request priority for the transit vehicle.

#### <span id="page-35-1"></span>**2.3.3. Transit detection**

Transit vehicle detection systems can be distinguished into four categories: driver actuated, point detectors, area detectors, and zone detectors([Baker et al.](#page-82-3), [2002](#page-82-3)). The first category *driver actuated* is desirable for vehicle detection. Preemption of emergency vehicles could be very beneficial; however, in a real-life study in Washington, experience learned that drivers of transit vehicles keep the transmitter on, even when priority was not needed or desired. *Point detectors*are used frequently to detect TSP. The detectors work well, but the limitations concern the limited information of the transit vehicles between the point detectors. *Area detectors* can monitor the movement of a vehicle through an area and, therefore, can be used to predict the arrivals of transit vehicles at intersections. The last category is *zone detectors*; these detectors know when a system is somewhere on the approach of an intersection (e.g. 150 meters) and is requesting priority.

#### <span id="page-35-2"></span>**2.3.4. Strategies for Real-time control and TSP**

Various strategies exist to implement adaptive/real-time traffic signal control with TSP. Although details of how each system implements adaptive control vary. In general, real-time signal control assesses the status of the network. With traffic flow predictions for cars, buses, and light-rail vehicles, traffic systems with TSP can modify signal timings to accommodate traffic demand efficiently. Real-time control requires more sophisticated detectors methods, communication, and processing capabilities than actuated signal control. Furthermore, when all modes are included in the real-time control, RTCS can optimize traffic in multiple ways – for example, the delay experienced by each vehicle or the delay experienced by each traffic participant. This is further explained in [section 2.4](#page-36-0).

TSP, in combination with real-time control, has been tested in real life, with critical performance criteria for transit identified as regularity and punctuality. For example, SCOOT bus priority, with inductive loop detection, proved to be effective and reliable. With more advanced information available, due to roadside beacons for the bus, TSP proved even more effective([Hounsell & Shrestha](#page-83-2), [2005](#page-83-2)). Systems that focus on special priority for modes other than the bus, such as a system with a particular priority for Light Rail Vehicles, have also been developed that control traffic in combination with a Genetic Algorithm formulation. The genetic algorithm optimizes traffic control for multimodal operations in
large urban networks. Solutions are constrained to preserve existing (and well-performing) Light Rail Transit predictive prioritystrategies in a Software-in-the-Loop environment ([A. Stevanovic, Kergaye, & Martin,](#page-85-0) [2009](#page-85-0); [A. Stevanovic et al.,](#page-85-1) [2015\)](#page-85-1). Basic signal timing for passenger cars is the most essential measure to optimize traffic signal control, compared to optimizing signal timing for vehicle transit or a combination. Additionally, this is the case when the objective criterion is for passenger cars, personal delay, or transit vehicle delay([J. Stevanovic](#page-85-2), [2011\)](#page-85-2).

Multiple real-time urban traffic control systems in literature have been developed to regulate global traffic, and extensions to these models have been made to include public transit priority in multimodal optimization. These multimodal realtime controlled systems will be described in the following section.

# **2.4. Multimodal real-time controlled systems**

Multimodal RTCS with multimodal optimization control all traffic by an objective function that optimizes all vehicle modes. While the traffic demand for private cars is still growing, the growing number of users of buses, light-rail vehicles, cyclists, and pedestrians, increases the complexity of finding the optimal signal control solution for an urban traffic environment [\(Christofa et al.](#page-82-0), [2013](#page-82-0); [Bhouri et al.](#page-82-1), [2015](#page-82-1); [Mein et al.,](#page-84-0) [2022\)](#page-84-0). Although signal control and multimodal priority systems have been widely deployed over the last few decades, these two systems often conflict with each other due to different control objectives [\(He et al.](#page-83-0), [2014](#page-83-0)). One of the identified issues in these systems is providing equity between modes. Another is estimating the vehicle delay, often based on existing real-time control systems.

In the following subsections, the used traffic detectors, control, queuing model, and objective function of multimodal real-time controlled systems are described.

### **2.4.1. Traffic sensors**

Multimodal RTCS use multiple traffic sensors, as not only passenger cars but also LRVs, buses, bicycles, and pedestrians must be identified. Which can be done with conventional traffic sensors, such as magnetic loop detectors, with upstream and downstream possibilities for passenger cars, but prescribes a detector for each traffic mode. Data-driven systems use more sophisticated laser or machine vision sensors. These sensors make it possible to identify multiple traffic modes with a single sensor, possible with LiDAR sensors [\(J. Wu et al.](#page-85-3), [2020;](#page-85-3) [Mein et al.](#page-84-0), [2022](#page-84-0)), or machine vision [\(Albiol et al.,](#page-82-2) [2011\)](#page-82-2). Additionally, queue lengths can be measured in real-time, or the traffic system can count the number of vehicles in the inner junction of an intersection [\(Boillot et al.](#page-82-3), [2006](#page-82-3)). However, these sensors are more prone to sensor faults when the weather is bad – for example, when it is dark or it is misty. Furthermore, less sophisticated are cheaper in terms of purchasing costs.

### **2.4.2. Traffic signal control**

In terms of utilized traffic signal control, three types are found. The first is a (traffic responsive) fixed timing control with transit signal priority [\(Yagar & Han,](#page-85-4) [1994;](#page-85-4) [Bhouri et al.,](#page-82-1) [2015\)](#page-82-1), here transit signal priority interrupts the proposed signal plan to provide priority for public transit. An alternative is adaptive traffic control; the traffic light switches dynamically when the option provides a better alternative than the current phase plan [\(Zeng et al.,](#page-86-0) [2019;](#page-86-0) [Christofa et al.](#page-82-4), [2016\)](#page-82-4), adhering to safety constraints but mainly focused on minimizing vehicle or personal delay, regardless of the type of road user. The last type is adaptive control, which uses a predefined signal set: a collection of phase plans is generated and optimized based on car delay using a genetic algorithm. This results in a collection of phase diagrams. During the prediction horizon of the multimodal RTCS, the adaptive control evaluates different green options for cars and public transit and chooses the plan that leads to the minimum or maximum of the objective criterion [\(He et al.](#page-83-0), [2014;](#page-83-0) [A. Stevanovic et al.](#page-85-1), [2015;](#page-85-1) [Mein et al.,](#page-84-0) [2022](#page-84-0)).

#### **2.4.3. Modelling**

For the estimation or prediction of traffic variables, older systems [\(Yagar & Han,](#page-85-4) [1994\)](#page-85-4) to newer methods([He et al.,](#page-83-0) [2014;](#page-83-0) [Xie et al.](#page-85-5), [2014;](#page-85-5) [Bhouri et al.,](#page-82-1) [2015;](#page-82-1) [Mein et al.,](#page-84-0) [2022\)](#page-84-0) use the delay to optimize multiple road users. For passenger cars, the delay is calculated according to a VQM. The VQM is easily implemented, and is computational fast. Yet, as previously mentioned, a VQM does not take spatial characteristics into account and delays are underestimated when queues are long. On the other hand, two identified systems apply shockwave theory to estimate delays or queue lengths, the system from [Zeng et al.](#page-86-0) ([2019\)](#page-86-0) which uses connected vehicle to estimate the number of vehicles in the queue. Secondly, the multimodal system from [Christofa et al.](#page-82-4) ([2016\)](#page-82-4) estimates car delay based on shock wave theory, comparable to the method in [subsection 2.1.2.](#page-27-0) However, the queue length is not considered explicitly. In these multimodal systems, queue model are only used for the delay calculation. The minimum of the delays is then used to find the optimal signal plan.

# **2.4.4. Optimization**

Multimodal RTCSs optimize based on either personal or vehicle delay. The personal delay is defined as the delay experienced by an individual road user, the vehicle delay is the experienced delay per vehicle. The researched multimodal system developed by [Christofa et al.](#page-82-4) ([2016\)](#page-82-4) made use of personal delay, and the articles of [He et al.](#page-83-0) [\(2014](#page-83-0)), [Mein et al.](#page-84-0) [\(2022\)](#page-84-0) use vehicle delay and include a vehicle priority list to give the possibility to provide additional priority for specific modes. [A. Stevanovic et al.](#page-85-1) ([2015\)](#page-85-1) tested multiple configurations and concluded that optimizing a predefined collection of signals plans based on vehicle-delay-optimization yields the best results, comparing this to transit-delay-optimization or a combination of transit- and vehicle-delay-optimization.

Using personal delay was found successful for objective optimization functions([J. Stevanovic,](#page-85-2) [2011\)](#page-85-2). Consequently, the higher the passenger occupancy of a transit vehicle, the higher the priority and the higher the benefit for transit users. However, when the car traffic demand is very high, personal or vehicle delay optimization converge to the same outcome. In addition, when transit vehicles such as LRVs have high occupancy, the reduction in personal delays grows until the system operates close to saturation; at this point, the optimization leads to the same solution as with a lower passenger occupancy. Providing additional priority for other traffic modes in such situations causes a lack of flexibility. Therefore, it is advised that experiments should incorporate different weights for various transportation modes applied in the person-delay optimization.

Using vehicle delay also has benefits with special priority for LRV. The priority is often defined as a user-specified list of priority weights. Public transit, such as LRVs or buses, receive a higher priority than passenger cars. This has the benefit that policymakers could include their decision-making goals directly in the priority weights of traffic modes. However, one of the identified issues in these systems is providing equity between traffic modes and the lack of flexibility in providing additional priority when the network demand grows during peak hours.

Providing additional priority also lacks in the traffic control of multimodal RTCS that use an objective function that optimizes all traffic modes. However, additional priority could be justified as spillback effects, such as spillover, can obstruct other traffic direction, or could lead to unsafe situations when right-of-way is given to a higher priority LRV.

# **2.5. Spillback control in MM real-time control systems**

Little research has been done on spillback modelling in multimodal real-time controlled systems. Regarding unimodal RTCS, [Noaeen et al.](#page-84-1) [\(2021](#page-84-1)) developed a control strategy that avoids and handles spillback effects, suggesting that more research is needed to include more modes and test transit priority. In the context of multimodal RTCS, it is suggested that a more sophisticated queue estimation model is used [He et al.](#page-83-0)([2014](#page-83-0)). [Christofa et al.\(2016](#page-82-4)) developed a multimodal RTCS with a more sophisticated queue model that applies shockwave theory but neglects the queue lengths in the traffic control and suggests that non-recurrent congestion, such as spillover and/or lane blockage should be included in the control strategies of multimodal real-time traffic control. A multimodal RTCS (SURTRAC) that detect and handles spillback is designed by [Xie et al.](#page-85-5) [\(2014](#page-85-5)). The system uses a vertical queuing model, and based on the state identification in this standing queue [\(Perez-Montesinos et al.](#page-84-2), [2011\)](#page-84-2), spillback can be detected and handled. However, SURTRAC is a patented multimodal RTCS([Xie et al.,](#page-85-6) [2015](#page-85-6)).

The results lead to insight about the effects on the performance of a multimodal RTCS. First, in terms of throughput, (transit) vehicles could achieve lower delays and a higher throughput when bottlenecks are solved sooner. Second, in control strategies where oversaturated traffic phenomena, such as spillover and lane blockage, are avoided or controlled, unsafe situations are avoided or solved faster so that less conflicts occur between different road users. For example, passenger cars spill over to an upstream intersection when an approaching light-rail vehicle crosses the street or when passenger cars halt pedestrians at the crosswalk due to downstream congestion.

The researched literature led to a conceptual framework that summarizes key aspects of researched articles. Namely, the included road users in the RTCS, whether priority is provided, if an objective function is included, the utilized traffic signal control, the traffic sensors, and the queue length estimation. Lastly, the spillback detection and control strategies of researched articles. A complete overview of the conceptual framework [\(Table 2.2](#page-38-0)) with mentioned contributions can

<span id="page-38-0"></span>Table 2.2: Literature overview (full table in [Appendix A](#page-87-0),[Table A.2\)](#page-88-0)



<mark>Legend.</mark> (brackets): Mentioned as possible extension of the model. Max: Maximize. Min: Minimize<br>**Road users**: PC (Passenger cars), B (Bus), T (Truck), LRV (Light-rail vehicle), BC (Bicycle), P (Pedestrian). Priority: BP ( Objective: TT (Travel time), TP (Throughput), D (Delay), QL (Queue length), CF (Common Flow), VD (Vehicle delay), PD (Personal delay).<br>Signal Control: FT (Fixed timing), Set (AC with signal plans predefined with genetic al

**Queue model**: S (Measured from simulation software), SWT (Shock wave theory), I/O (Input/output method), CTM (Cell transmission model), CV (Connected vehicles), VQM (Vertical queuing model)

**SB type**: LB (lane-blocking), SO (Spillover). Detection: LC (Link capacity), OSI (Oversaturation Severity Index), QT (Queue Threshold),

Prob() (Probability of spillback type), SD (Speed detection).

**SB control**: OF (Objective function), PinOF (Penalty in objective function), M- (Metering from US and/or DS direction), US (Upstream), DS (Downstream), SS (Sidestreet)

be found in [Table A.2](#page-88-0).

# **2.6. Conclusion**

In the literature, control strategies have been developed that detect or control ocurring spillback effects. In this chapter the aim was to answer the first two subquestions of RQ1: *How could spillback be detected?* And *how could spillback be avoided or handled?*. First, the findings of current research is described, after that, implications for future research are made

#### Findings

It was found that the spillback detection approach was dependent on the utilized traffic sensors, and whether a queue method was used for detection. Two types of spillback are detected in literature, namely, lane blockage or spill over to an upstream intersection.

A commonly used approach in literature for spillback detection was using the link capacity of a lane. When the capacity was defined in terms of the number of vehicles, the traffic system counts the number of incoming vehicles, the number of vehicles are then compared against the link capacity, assuming a fixed vehicle length and vehicle gap. This method is often paired with a VQM. When the capacity was defined in terms of meters, the queue length was estimated in realtime using a queue model. For the queue length estimation, in most cases, shock wave theory was applied. Also, in some cases the queue length was used in spillback detection, but it was retrieved directly from the test environment. For both approaches, when the queue exceeds the link threshold, spillback is detected. Other cases that detect spillback, include a speed threshold, which detects spillback when the speed is low enough at the location an advance detector, another identification of the type of spillback, namely, the time a vehicle spends in the queue, and the queue length determined whether spillback is sustained or should be dealt with. The most pragmatic approach was the use of sophisticated traffic sensors that can supervise if spillback occurs at an inner junction, or when lane blockage occurs. This method performs well in good weather conditions.

Regarding spillback control strategies, mainly two approaches were found. The first approach was a form of metering by limiting the inflow with a green time reduction for critical phases at upstream intersections, promoting the traffic flow with green time extension for the congested link, or both. The second approach is with the optimization of an objective function. The optimization minimizes queue length, or maximizes of the congested link or, to solve spillback queues.

In terms of avoiding spillback, objective functions were used to find an overall optimal flow of the intersection, or to find the maximum average throughput for the whole intersection.

The researched literature led to a conceptual framework that summarizes key aspects of researched articles, depicted in [Table A.2](#page-88-0). Namely, the included road users in the RTCS, whether priority is provided, if an objective function is included, the utilized traffic signal control, parameters for the spillback detection method, and control strategies of researched articles.

#### Implications for future research

Regarding RTCS with a multimodal optimization algorithm, spillback detection and control strategies are almost never included in the traffic light control. The systems often utilize a VQM for delay estimation, therefore queues are not incorporated explicitly. As spillback detection is almost never incorporated in the system control, little research has been done on the effects of spillback control strategies in a multimodal RTCS on the traffic and network performance.

Insight is given for spillback control strategies, whether benefits are found for traffic participants, such as lower delays for congested links and the delay performance for the whole intersection. Secondly, when spillback effects, such as spillover and lane blockage, are handled, this leads to less direct conflicts between different road users which decreases unsafe situations.

This is realized by developing a spillback control strategy when spillback is detected. In research there was found that in many approaches the queue length was used for spillback detection, furthermore, when the queue length was measured almost always shock wave theory was applied. This comes with that advantage that for the calculation of delay, the constructed shock wave profile can also be used for a more accurate calculation for incoming traffic. Therefore, a queue method will be designed for the detection of spillback. The approach for the control strategy there will be looked at what can be integrated in the system under study. The description of this system is described in the next section.

# 3

# <span id="page-40-1"></span>MobiMaestro-Flow – System under study

The utilized traffic system for the thesis is MobiMaestro-Flow, a multimodal real-time controlled traffic system. MM-Flow continuously optimizes individual intersections by minimizing the waiting costs, which consist of the multiplications of the waiting time per vehicle, the movement, and object priority. Furthermore, upstream and downstream intersections exchange information so the optimization algorithm can consider this in successive optimization iterations. The utilized algorithm is referred to as 'TFE'.

Individual intersections are made up of agents exhibiting intelligent behaviour and a drive to cooperate, as proposed in [Van Katwijk](#page-85-9) [\(2008](#page-85-9)). The optimization approach is multimodal, meaning that the traffic light control optimization includes multiple modes. Objects and movements can be assigned different priorities so that TFE can favour traffic objects in green time allocation. Furthermore, V2I technologies can provide preemption so that an emergency vehicle receives a green wave in emergency [\(He et al.,](#page-83-0) [2014](#page-83-0)). Traffic flow predictions are made by free-flow travel time between an upstream intersection to the stopbar of an intersection, and predictions have a planning horizon of 120 seconds and an update frequency of 1 to 5 seconds. The following section describes the architecture. After that, the utilized traffic sensors, the operated traffic signal control, the queuing method, and the optimization algorithm are described.

<span id="page-40-0"></span>

Figure 3.1: The logical structure from TFE (adapted from [Mein et al.](#page-84-0) [\(2022\)](#page-84-0))

# <span id="page-40-2"></span>**3.1. Architecture**

The architecture from MobiMaestro flow consists of five levels, in which the first level is the system's input from the traffic sensor. The corresponding traffic sensor modules handle and process the information in level two. The module processes the data into a suitable format for the traffic flow engine (TFE). Data delivered to the TFE are object data from vehicles travelling on the road, platform and LRV-states, and infrastructure states, such as traffic light signals. The traffic

flow engine creates crossing requests for every mode and predicts traffic flows, serving as input for the optimization in levels 3 and 4. Level 5, the policy configuration, is also an input for the optimization. The policy configuration provides additional priority for particular traffic objects or movements in the network. In level 3, local intersections are optimized using the objective function. In level 4, the corridor optimization, the traffic flow predictions are also included in the optimization. An overview of the logical structure is depicted in [Figure 3.1.](#page-40-0)

# **3.2. Traffic sensors**

The traffic sensors provide input so the traffic flow engine can digitally reconstruct cars, public transit, cyclists, and pedestrians. Hardware and software modules classify objects, determine vehicle speeds, positions, and vehicle lengths, and are placed upstream and downstream of the intersection. In a previous test case in San Francisco, the utilized traffic sensors were LiDAR sensors; however, other traffic sensors, such as video detection units, are also applicable. In the case of LiDAR sensors, Qortex software transforms the point cloud data from LiDARs into real-time object information.

Regarding public transit, additional sensors are placed on public transit platforms to measure the number of waiting passengers and predict dwell times and departure times of Light-rail vehicles. The traffic lights utilize D4 signal controllers for the real-time management of traffic signal lights. Here, NTCIP retrieves controller status and actual phase information and sends desired plans. Generated phase plans are bound to safety constraints, such as clearance time, minimum green time, and local rules, such as the phase structure used—the next section elaborates on the traffic signal control in MobiMaestro-Flow.

# <span id="page-41-1"></span>**3.3. Traffic signal control**

#### **3.3.1. Phase structure**

The traffic signal control uses a ring-and-barrier structure, following the NEMA phasing convention, meaning that at a signalized intersection, such as the one depicted in [Figure 3.2a](#page-41-0), eight phases are used. These phases can be used by various users, such as cars, public transit, and pedestrians. Therefore, pedestrian phases P2, P4, P6, and P8 can receive green simultaneously as phases 2, 4, 6, and 8, respectively.Furthermore, turning right follows the same phase as the straight – for example, phase 8 for the west side also serves the right turn for that direction. The intersection in [Figure 3.2a](#page-41-0) leads to the conflict matrix in [Figure 3.2b](#page-41-0).

<span id="page-41-0"></span>



(a) Phases at intersection

(b) Conflict matrix car phases

Figure 3.2: Phases and resulting conflict matrix profiles at the intersection

# **3.3.2. Ring-barrier diagram**

The ring-and-barrier structure groups phases in a continuous 'ring' and separates the conflicting traffic phases by making them sequential in a ring or by adding a barrier between the movements([Office of Operations,](#page-84-6) [2021](#page-84-6)). Multiple configurations of signal plans are generated, in which separate rings in the same barrier can receive green similarly; otherwise, the directions are conflicting. Following the conflict matrix in [Figure 3.2b](#page-41-0), this could lead to the phase diagram depicted in [Figure 3.3.](#page-42-0) This phase diagram is similar to the phasing configuration of the intersection of 3rd Street & 16th Street in San Francisco. The phase plans are generated based on minimizing the waiting costs, calculated by the sum of crossing and virtual crossing requests. Among other things, these requests depend on the time the vehicle waits at the traffic light, which the queuing model determines.

<span id="page-42-0"></span>

Figure 3.3: Example of phase diagram, similar to configuration of 3rd St. & 16th St.

# **3.4. Modelling**

The prediction of traffic flows and the calculation of delays are used in the crossing requests. First, the elements of crossing requests are explained, after that, the delay alculation is explained.

### <span id="page-42-1"></span>**3.4.1. Crossing requests**

All possible movements and intersection crossings are described using crossing requests. A crossing request is defined as any traffic object *n* with the intention of crossing the road at an intersection. Traffic objects can be any road user, such as an LRV, passenger cars, buses, cyclists, and pedestrians, but also an emergency vehicle, such as a firetruck. These requests are weighted and are determined with the following function:

Crossing request(*i*) = Waiting-time(*i*)<sup>*Y*</sup> · (
$$
\alpha
$$
 · object-priority) · ( $\beta$  · movement-priority) (3.1)

The *Waiting-time* represents the red-light delay of the traffic object, which is measured as the time stopped in front of a red light in seconds. The *object-priority* is a preset priority weight assigned to traffic objects to accommodate additional for specific modes, such as a higher priority for the Light Rail Vehicle to promote transit priority. *Movement-priority* is a preset priority assigned to the road network to configure the additional priority for directions. Parameters, such as  $\alpha$ ,  $\beta$ , and  $\gamma$ , control the impact of individual terms of the formula. The *Waiting-time* is set to the power of  $\gamma$  to give an exponential penalty for waiting costs. Consequently, when a vehicle waits before a red-light, the penalty for the waiting time grows exponentially after three seconds, so that long waiting times are avoided. For the experiments performed in San Francisco and this thesis,  $\gamma$  was configured as 1, therefore, it is left out in further descriptions and equations. Furthermore, starvation prevention is implemented, which accounts for possible sensor faults. When a phase is skipped for multiple cycles and has not received a green light status for three minutes, a crossing request is made for that direction with a high priority to establish a green phase in the next minute.

For vehicles approaching the intersection, a virtual crossing request is made for the downstream direction of the intersec-

tion. The reason for a virtual crossing request is that it is less accounted for in the optimization than an actual crossing request, as the outcome of the prediction is not certain. The structure of virtual crossing requests is similar to the standard crossing requests. However, the request is multiplied by the probability of a turn ratio based on historical traffic flow data and parameter  $\theta$  so that the virtual crossing request is assigned fewer costs than regular crossing requests.

Virtual crossing request(*i*) = Waiting-time(*i*)  $\cdot$  ( $\alpha$   $\cdot$  object-priority)  $\cdot$  ( $\beta$   $\cdot$  movement-priority)  $\cdot$  ( $\theta$  $\cdot$  turn-probability) (3.2)

#### **3.4.2. Delay calculation**

For the optimization algorithm, the arrival time of vehicles is predicted, and the delay is measured as the time stopped in front of a red light in seconds. Traffic flow arrivals are estimated by the free-flow travel time between an upstream intersection to the stopbar of an intersection, in a similar way as described in [subsection 2.1.2](#page-27-0). For the arrival prediction, the time instant of the upstream detector is used, in which the distance of the upstream detector, divided by the freeflow speed resemble the travel time from the upstream detector to the stop-bar, as shown in [Equation 3.3.](#page-43-0) The departure time for the first vehicle is equal to the time instant the traffic light turns green,  $T_g$ , plus the start-up delay. For successive vehicles, the departure time is dependent on the configured headway *H*. In the case of MM-flow, this is configured as two seconds, meaning that every two seconds a vehicle discharges when the light is green.

<span id="page-43-0"></span>
$$
A(i) = \text{Upstream detector arrival}(i) + \text{distance of upstream-detection}/v_{freeflow}
$$
\n(3.3)

The delay is also calculated using the expected arrival and departure time. Vehicles are expected to depart according to an assumed vehicle headway of two seconds. The total waiting time is then equal to the number of vehicles, multiplied by the headway, and summed with the waiting time of the first vehicle. The waiting time calculation for the first vehicle is dependent on the expected arrival departure time, and the start-up delay, as shown in [Equation 3.5.](#page-43-1) The departure time is equal to the time instant the vehicle receives green,  $T_g$ .

<span id="page-43-2"></span>Total Waiting time of queue = Waiting time (1) + 
$$
\sum_{i=2}^{I} (i-1) \cdot H
$$
 (3.4)

<span id="page-43-1"></span>
$$
Waiting time(1) = Startup-delay + (D(1) - A(1))
$$
\n(3.5)

$$
D(1) = T_q \tag{3.6}
$$

[Figure 3.4](#page-44-0) depicts a numerical example: the waiting time of the second vehicle in the queue is Waitingtime(2) =  $H$  + Waiting time(1) = 33*s*, assuming a headway of 2 seconds and a waiting time of 31 seconds for the previous vehicle. The waiting time of a car can be calculated according to [Equation 3.4](#page-43-2).

For longer queues, more uncertainty could occur compared to a queue that consists of only three cars. The delay computation are efficient; however, in congested cases, the queue discharge rate is overestimated. The spatial queue length is not considered in the queue discharge and is discharged assuming a fixed headway *H* between the vehicles. This way, the cars are modelled leaving the queue for every *H* seconds.

The calculated waiting times for vehicles are used in the TFE for finding a minimum delay for all approaches with an accompanying signal plan. This will be explained in the next section.

# **3.5. Optimization algorithm**

The TFE continuously minimizes the total weight of crossing requests of all road users, given the waiting time and the preconfigured mode and movement priority. With the national rules and possible phase timing, a predefined set of phase timings are generated using a genetic algorithm, in which the basic signal timings are optimized. The generated set of phase diagrams results in a collection of 20.000 phase diagrams. The height of the waiting costs determine which direction is allocated green time, in which the waiting costs is the result from the sum of crossing and virtual crossing requests.

<span id="page-44-0"></span>

Figure 3.4: Example of waiting time calculations

#### **3.5.1. Local intersections costs**

When in operation, the TFE creates and updates crossing requests for existing and entering traffic objects*i*. The waiting costs per direction at local intersections are the sum of crossing requests.

Waiting costs of actual approaches 
$$
=\sum_{i=1}^{I}
$$
 Crossing request(*i*) (3.7)

With every predicted change in crossing requests, waiting costs are calculated by evaluating phase timings per second, constrained by the possible and applicable phase timings. The crossing requests per traffic object *i* are summed for the total number of vehicles *I* in the approach.

#### **3.5.2. Network costs**

When controlling multiple intersections, TFE generates virtual crossing requests for predicted downstream approaches. When a vehicle approaches an intersection, the prediction is made that the vehicle will continue downstream to the next intersection. Consequently, for a network consisting of multiple intersections, the waiting costs are determined by the sum of crossing requests per approach for vehicles *i* and the sum of virtual crossing requests for vehicles *i*. The total waiting costs are calculated for every phase and intersection, which is indicated in the next subsection.

Total waiting costs = 
$$
\sum_{i=1}^{I} \text{Crossing request}(i) + \text{Virtual crossing request}(i)
$$
 (3.8)

#### **3.5.3. Optimization**

Every evaluation loop results in waiting costs for every crossing request at every intersection. The objective function is sum of the total waiting costs per intersections. The goal is to minimize the overall waiting costs per intersection over a planning horizon of 120 seconds with an update frequency every 1 to 5 seconds. To achieve this, a collection of predefined phase diagrams are generated, with a varying sequence of phases, varying green times, and the possibility of phase skipping, leading to a total of 20.000 phase diagram per intersection.

Set of phases

\nminimize:

\n
$$
\sum_{\text{phase}}^{\text{Set of phases}}
$$
\nTotal waiting costs(phase, intersection)

\nsubject to: Phase structure

\nSafety constraints

\n(3.9)

The optimization algorithm evaluates each phase diagram with the resulting waiting cost over the period of the planning horizon. The optimization is subject to local regulations, such as the used phase structure, and safety constraints, such as the red time clearance. The phase structure entails that some phases can receive green simultaneously, such as when phase 8 is allocated green time, phases 3 or 4 can also receive green (see [Figure 3.2b](#page-41-0)), dependent on the height of the waiting costs one of both is allocated green time. The result of the optimization leads to a designated signal diagram.

#### **3.5.4. Policy configuration**

The result of waiting costs also depend on the configured priorities in the algorithm, as explained in [subsection 3.4.1](#page-42-1). With traffic object priority and movement priority configuration, policy goals or ambitions, such as improving public transit by improving punctuality, can be implemented using a higher priority weight for public transit vehicles. Consequently, public transit, such as an LRV, receives more priority in green time allocation than a car. On the other hand, specific routes in the corridor can receive different priorities – for example, to promote the traffic flow on the arterial, routes on the arterial are assigned a higher priority than sidestreets.

[Table 3.1](#page-45-0) shows an example configuration used by the San Francisco case study. Here, the LRV was assigned a priority of 2,000, whereas cars were given a priority of 5. Additionally, an emergency vehicle is given a significantly higher weight than other vehicles (100.000) to allocate a green wave for a priority vehicle, such as a fire truck, in the case of an emergency.



<span id="page-45-0"></span>Table 3.1: Example of mode prioritization in San Francisco case study

To research the effects on traffic performance and network performance, a spillback detection method and control strategies are designed. The additional components are explained in the next chapter.

# 4

# Design of spillback component

The previous chapter described the architecture and components of MobiMaestro-Flow. This chapter proposes the spillback component that extends the architecture. This way, spillback effects, such as lane blockage and spillover, can be detected, furthermore, a spillback control strategy so that the identified spillback effect is handled.

# **4.1. Spillback detection**

In the literature was concluded that the spillback detection (and control strategy) was dependent on the utilized traffic sensors. In the next sections, first, the traffic sensors needed for the spillback detection method are described, and after that, the queue method for queue length estimation, lastly the spillback detection method.

## **4.1.1. Traffic sensors**

MobiMaestro-Flow, described in the previous chapter uses sensors that detect lane areas and vehicle speed and location, comparable to a traffic sensor that uses machine vision or laser sensors. The advantage of these sensors is that the queue length can be measured directly, such as with the method of [J. Wu et al.](#page-85-3) ([2020\)](#page-85-3). However, these sensors are vulnerable to adverse weather conditions and the sensors are far more expensive than magnetic loop detectors. Furthermore, approaches in the literature are also applicable when these sensors are used. Therefore, it is chosen to use magnetic loop detectors for spillback detection.

### **4.1.2. Detection with lane threshold**

In combination with the chosen sensors, a threshold determines whether spillback is detected. The threshold is the length of the link (*llink*) or the length of a turning bay (*lturningbay*). Spillback is detected when the estimated maximum queue length ( $L_{max}^n$ ) during a cycle  $n$  exceeds the queue threshold. Which leads to the following conditions:

> Spillover detection:  $L_{max}^n \ge l_{link}$ Lane blockage detection:  $L_{max}^n \ge l_{turningbay}$

Consequently, spillover is detected when the queue length exceeds a link length in which the vehicles spill over to an upstream intersection. Furthermore, lane blockage is detected when the queue length is exceeds the turning bay length. When one of these conditions does not hold, spillback is communicated as true to the RTCS. The queue length is estimated according to a queue definition and a queue model, explained in the next sections.

### **4.1.3. Queuing model for estimation of maximum queue length**

The maximum queue length during a cycle is estimated for spillback detection. This can be done using the approach by [H. Liu et al.](#page-83-3) [\(2009](#page-83-3)) that describes the queue dynamics at congested intersections. The LWR shockwave theory from [Lighthill & Whitham](#page-83-4) [\(1955](#page-83-4)) is applied to estimate queues and the cycle-based method estimates maximum (and





residual) queue length in real-time at the end of each traffic light cycle. The approach assumed that event-based traffic signal data is available for identifying traffic state changes, distinguishing vehicle queue discharge flows from upstream traffic arrivals so that even long queues can be estimated [\(H. Liu et al.,](#page-83-3) [2009](#page-83-3)).

#### Queue definition

A queue consists of a head and a tail, the head being the first vehicle at the stopbar. The tail is the last vehicle in the queue. It is chosen to include a vehicle in the queue length when an arriving car drives slower than 18 km/h (5 m/s). This value is chosen, as in the test environment, 5 m/s was a speed where the vehicle almost always ended in a total standstill. For higher values, the car was part of heavy traffic and started accelerating again after a period.

#### Fundamental diagram

Several assumptions are made in the real-time estimation of the maximum queue length. In applying shockwave theory, the derivation of shockwave speeds uses the relations between macroscopic flow characteristics, called fundamental diagrams, as a basis. The variables density *k*, speed *u*, and flow *q*, represent the same information. Density is defined as the number of vehicles per unit of distance, flow is defined as the number of vehicles per unit of time, and speed is measured in the distance per unit of time. These three variables are referred to as traffic states, where vehicles can transition from uncongested to congested states. This can be summarized in four points indicated in a fundamental diagram and will be used in the queue length estimation [\(Lighthill & Whitham,](#page-83-4) [1955](#page-83-4); [Hoogendoorn & Knoop,](#page-83-5) [2016a](#page-83-5),[b\)](#page-83-6).

On the left side of [Figure 4.2,](#page-48-0) these points are schematized and represent the traffic states occurring at signalized intersections.

 $\mathbf{State}\left(\mathbf{1}\right)$  Inflow  $q_i$ , represents the inflow of the traffic at point  $(q_i,k_i).$ 

**State (2)** Jam density *k<sup>j</sup>* , represents the density in which the traffic has come to a full standstill with flow 0, at point  $(0, k_i)$ .

**State (3)** represents an empty road, at point (0*,* 0).

**State (4)** The maximum flow *qmax* represents the maximum flow of the traffic, where the critical density is the density where the traffic flow switches from the maximum flow to a decreased flow. The maximum flow with the critical density is at point (*qmax, kc*)

There is assumed that vehicles act according to the fundamental diagram, and depart with the maximum possible flow

<span id="page-48-0"></span>

(*qmax* when possible. For example, a traffic light turns green after being red for an extended period.

Figure 4.2: Critical points in FD (left), and shockwave speeds in FD (right)

#### Basic shockwave profile

With a known fundamental diagram of the used road, signalized intersections can apply SWT to estimate queue lengths and propagation. In [Figure 4.3](#page-49-0) the estimated traffic states are shown, corresponding to the traffic states in the fundamental diagram. For example, if the traffic state changes between 1 and 2, the tangent of the line is equal to  $\omega_3$  from the fundamental diagram [\(Figure 4.2](#page-48-0)). The shockwave speed is calculated according to the formulas [Equation 4.1](#page-48-1) - [Equation 4.4.](#page-48-2)

<span id="page-48-1"></span>
$$
\omega_1 = v_{free} = \frac{q_i^n - 0}{k_i^n - 0} = \frac{q_i^n}{k_i}
$$
\n(4.1)

$$
\omega_2 = \frac{q_{max} - q_i^n}{k_c - k_i^n} \tag{4.2}
$$

$$
\omega_3 = -\frac{q_i^n}{k_j^n - k_i^n} \tag{4.3}
$$

<span id="page-48-2"></span>
$$
\omega_4 = -\frac{q_{max}}{k_j - k_c} = \tag{4.4}
$$

For the maximum queue length estimation, other variables, such as the green time  $(T_g^n)$  and red time  $(T_r^n)$  are needed to estimate the maximum queue length ( $L^n_{max}$ ) and the time it occurs ( $\vec{T^n_{max}}$ ), during cycle  $n.$  But also variables from the advance detector are needed, such as the time instant the queue builds up to the advance detector  $(T_A)$ , the time instant the queue starts propagating forward  $(T_B)$ , and the time instant the queue is discharging  $(T_C)$ . All these variables are depicted in [Figure 4.3.](#page-49-0)

#### Maximum queue length estimation

The shockwave speeds in combination with detector data (points A,B, and C) are used to estimate the maximum queue length ( $L_{max}^n$ ) and when it occurs ( $T_{max}^n$ ). As was indicated above, between the diamond breakpoints  $T_B$  and  $\hat{T_C}$ , the traffic state is at maximum flow:  $(q_{max},k_c)$ . State  $(q_i,k_i)$  is after  $T_C$  but before the next traffic light cycle  $(T_r^{n+1})$ . With  $\omega_2$  and  $\omega_4$  known, time instances  $T_B$  and  $T_C$ , and the detector length  $L_d$  can be used to estimate  $L_{max}^n$  and *T n max*:

$$
L_{max}^n = L_d + (T_C - T_B) / \left(\frac{1}{\omega_2} + \frac{1}{\omega_4}\right)
$$
 (4.5)

$$
T_{max}^n = T_B + (L_{max}^n - L_d) / \omega_4
$$
 (4.6)

The current example assumes a homogeneous road and a constant inflow. However, the traffic variables density, flow and speed can also be estimated with event-based data from loop detectors, furthermore, a residual queue can also be estimated. As designing a queue model is not the main focus of the thesis, this is explained in [Appendix A](#page-87-0). Nonetheless, the outcome of the queue length is compared against the lane threshold, in which the outcome is that spillback is detected or not. In the next section, it is explained what strategies are developed to control the identified spillback effects.

<span id="page-49-0"></span>

Figure 4.3: Applied SWT at signalized intersection

# **4.2. Spillback control strategies**

Two spillback control strategies are designed, and are depicted in the logical structure of the TFE in [Figure 4.4](#page-49-1). Components that already are implemented in the TFE are blurred to highlight the additional spillback components. The main difference between the two control strategies, is that the first strategy alters the optimization outcome, and the second strategy influences the optimization outcome. The strategies are explained more in detail in the next sections.

<span id="page-49-1"></span>



### **4.2.1. Control strategy 1: override**

The first spillback control strategy is a rather abrupt method that enforces the congested link to a maximum amount of green time so that when spillback is detected it is solved as soon as possible. The strategy has a high advantage for the congested direction; however, other directions may suffer unnecessary delays. In [Figure 4.4,](#page-49-1) the approach is depicted, the queue length is estimated for every cycle and communicated to the spillback detector. The spillback detectors compare

the queue length to the queue threshold, and when the threshold is exceeded, spillback is communicated to the crossing requests generator. The crossing requests generator consists of the existing LRV CR Generator, Vehicle CR Generated, and Pedestrians CR generator (see [Figure 3.1\)](#page-40-0). In this strategy, it is communicated to the intersection optimizer that spillback is detected, but no additional calculations are done.

The intersection optimizer generates the phase diagrams as usual (as explained in [chapter 3\)](#page-40-1). Just before the generated phase diagrams are communicated to the NTCIP Module, the spillback direction of a phase, which is the direction in which the spillback effect occurs, is set to green and conflicting directions are set to red. The green time of the spillback direction is continued until it is communicated that no spillback is detected anymore. In this strategy, the phase constraints of the system, such as the phase structure and safety constraints are not taken into account, meaning that the order of ring-and-barrier structure is neglected, and safety constraints, such as the minimum green time of non-spillback directions. The strategy illustrates a theoretical gain, in terms of a proactive spillback control approach.

# **4.2.2. Spillback strategy 2: penalty factor**

The second spillback control strategy penalizes the direction affected by spillback so that when spillback is detected, the direction receives higher priority, compared to other directions. This leads to an earlier green and advantages for the congested direction; however, less absolute priority is given compared to the 'override' strategy, and fewer unnecessary delays occur for disadvantaged directions. The spillback penalty factor induces a trade-off between higher prioritized public transit or a direction affected by spillback. Furthermore, a trade-off is induced for a higher prioritized spillback direction and possibly longer waiting cars in a disadvantaged direction.

For every cycle, the queue length is estimated, and the queue length is communicated to the spillback detector. The spillback detector compares the queue length to the queue threshold, and spillback is detected when the length exceeds the threshold. Then, a binary variable  $\epsilon$  is passed as a 1 to the crossing requests generator. The crossing requests generator consists of the existing LRV CR Generator, Vehicle CR Generated, and Pedestrians CR generator (see [Figure 3.1](#page-40-0)). In this strategy, additional priority is given to spillback directions in the crossing requests. This leads to an alternative calculation:

Crossing request(*i*) = Waiting-time(*i*)  $\cdot$  ( $\alpha$   $\cdot$  object-priority)  $\cdot$  ( $\beta$   $\cdot$  movement-priority)  $\cdot$  spillback-priority<sup> $\epsilon$ </sup> (4.7)

The generated crossing requests are communicated to the intersection optimizer, which optimizes the traffic as usual. The generated phase diagrams are communicated to the NTCIP module.

In the next chapter, it is described how the designed spillback detection method and control strategies are implemented in the test environment.

# 5

# Implementation of spillback component

# **5.1. Multimodal RTCS test framework**

<span id="page-51-0"></span>In [section 3.1,](#page-40-2) the architecture of MobiMaestro-Flow has been described. The test framework is a Software-In-The-Loop simulation (SILS). A SILS consists of a microscopic simulation, a virtual traffic controller running on the same computer, and a communication interface for information exchange between the microscopic simulation and the virtual controller([A. Stevanovic, Abdel-Rahim, et al.](#page-85-10), [2009\)](#page-85-10). The overview of the SILS is depicted in [Figure 5.1.](#page-51-0)



Figure 5.1: Test framework SILS

The used simulation environment is SUMO [\(Lopez et al.,](#page-83-7) [2018\)](#page-83-7), an open-source traffic simulation environment capable of executing microsimulations with individual vehicle behaviour. A demand profile and turning ratios are configured to run scenarios with varying traffic demand; in addition, traffic light control constraints provide the (legal) framework in which the traffic light optimization may optimize. Vehicle routes are generated using the given traffic volumes and turning ratios. Other configurable parameters are the weights of the vehicle priority and the route priority and all are used in the optimization of the virtual traffic controller 'TFE'.

The TFE generates crossing requests and, based on those crossing requests, optimizes the intersections, as is explained in [chapter 3.](#page-40-1) The communication interface retrieves traffic states from SUMO using the SUMO Traffic Control Interface (TraCI). For example, the interface retrieves the vehicle speed, length, and position using placed traffic sensors; in addition, traffic and traffic light states from SUMO, such as the green times. The communication interface then communicates the values to the virtual controller using the MQTT-protocol([Light](#page-83-8), [2017\)](#page-83-8). The types of messages sent and received are traffic light states, vehicle locations, and requested and current traffic light phase states. In the TFE

controller, the optimal signal control is calculated and communicated back to the communication interface, which then sets the new traffic light states in SUMO using TraCI. The test framework provides an environment to test traffic experiments and based on these experiments, generates output. The output data is then used for analysis purposes to describe effects of configured methods on the traffic and network performance.

In the test framework a spillback detection method is added, and two spillback control strategies. These components are described in the following sections.

# **5.2. Spillback detection**

A queue threshold detects spillback effects, a common approach among the researched literature([Wang et al.,](#page-85-11) [2016;](#page-85-11) [Noaeen et al.](#page-84-1), [2021;](#page-84-1) [Mohajerpoor & Cai](#page-84-4), [2020](#page-84-4); [H. Zhang et al.](#page-86-2), [2020\)](#page-86-2). In the framework, the estimated queue length is compared to the lane threshold of the traffic network. When the queue length exceeds the lane threshold, spillback is detected, for example, when the queue length exceeds the turning bay length or traffic spills over to the upstream intersection.



Figure 5.2: Queue length estimation with spillback detection

In the communication interface, the spillback detector is called, the spillback detector runs every timestep for every intersection-lane. In the function, the queue length estimation function is called which returns the queue length in meters, the front-lane id, which lane has the longest queue, and the total lane length. The queue length is compared to the lane threshold, which is similar to the lane length. If the queue exceeds the threshold, *isSpillbackDetected* is set to True (or 1) for that lane. Otherwise, the value of that *isSpillbackDetected* is False (or 0).

After that, every phase is coupled with the value of *isSpillbackDetected* and processed. The result is an intersection Id with a list of phases and a true or false related to the spillback state. Depending on the spillback control strategy, the result is processed further.

# **5.2.1. Queuing method**

The queues are retrieved from the SUMO environment. The reason for this is to narrow the research scope, a queuing model requires calibration and verification to work correctly. The operation of the queue model is important, as the queue length is used for the spillback detection. When the queue length is overestimated spillback could be perceived as true when it is not the case. On the other hand, when the queue is underestimated spillback can be perceived as false when in reality spillback occurs. The queue function calculated the queue for every time step for every lane in all intersection approaches. SUMO distinguishes lanes and edges: an edge can consist of multiple lanes. An intersection approach often exists of multiple consecutive edges.

The algorithm's input is the id-number of an intersection, and is needed to retrieve the lane Ids so that other variables can be retrieved. For every timestep, the vehicles currently on the lane, the vehicle speeds, vehicle lengths, and the vehicle positions are retrieved. The vehicle positions are retrieved using the SUMO function *getLanePosition(vehicle-Id)*. As the outcome of this function is the beginning of the link approach to the vehicle's front bumper. To actual queue length is calculated by the sum of the lane position and the vehicle length is subtracted from the lane length. After that, the partial queue of the lane is calculated by iterating through the list of current vehicles and comparing the vehicle speed to

a speed threshold. After calibration, it was found that 18 km/h (5 m/s) was a value where vehicles were nearing a halt in the simulation due to queues, instead of a temporary deceleration. The tail of the queue is identified as the lane length minus the furthest car plus the vehicle length.

The queue length is calculated for every lane. However, when no queue detected is on the front lane, no queues are found for the other lanes. When the lane is downstream, first is checked whether the front lane is saturated. When this is the case, the queue length is the sum of the lane queues. The algorithm returns the queue length for every intersection direction in meters, the front intersection lane with the longest queue (phase), and the lane threshold's total length.

**Algorithm 1** Queue length estimation

```
Retrieve all lanes of intersection
for lane do
  Get:
  VEHICLE: Id, speed, length, position.
  LANE: Id, length.
  Calculate queue:
  if vehicle speed < speed threshold then
    queue = lane length - min(vehicle position) + vehicle lengthend if
  if front-lane saturated then
    Total queue = front-lane queue + upstream-queue
  else
    Total queue = front-lane queue
  end if
end for
return total queue, Id of congested front-lane, length of front-lane
```
The spillback detector uses this output to compare the total queue to the queue threshold (length of the fron-lane). If the queue threshold is exceeded, the Id of the congested front-lane is communicated to the TFE. The TFE then applies a spillback control strategy for detected spillback.

# **5.3. Spillback control strategies**

Two types of strategies were implemented. The first is an override function, which immediately sets the congested lane to green. The second strategy includes a spillback component in the optimization algorithm so that a spillback direction receives a higher priority and receives green earlier, compared to the situation without the spillback component.

### **5.3.1. Spillback strategy 1: override**

In the control strategy is override, the spillback detector retrieves whether spillback is detected. When this occurs, the overflowing phases are communicated to the TFE([Figure 5.3](#page-54-0)). The TFE generates a collection of phase diagrams for optimal traffic signal control, after that, the phase diagrams of the intersections with spillback effects are overridden. The spillback direction is allocated green time, until the spillback state is controlled, meaning that it is communicated that no spillback is detected anymore. Furthermore, during this process, conflicting directions are allocated red time. When the spillback effects are solved, the traffic light control follows the conventional signal plan again. It should be mentioned that this strategy affects the proposed phase plan and traffic light constraints, such as the phase structure and safety constraints, are not taken into account. Therefore, the approach does not lead to signal plans that are valid in real life.

<span id="page-54-0"></span>

Figure 5.3: Congested lane is set to green

# **5.3.2. Spillback strategy 2: Penalty factor in objective**

<span id="page-54-1"></span>In the control strategy with a penalty factor in the objective function([Figure 5.4\)](#page-54-1), the spillback states identified by the spillback detector are messaged through the MQTT-protocol. The intersection optimizer processes the received data.



Figure 5.4: When spillback is detected, a penalty factor is activated in the saturated lane

#### **Algorithm 2** Spillback penalty algorithm

```
Objective min(waiting costs)
Waiting cost calculation for every movement:
if queue length > lane length (spillback) then
  return Optimization result ⋅ Penalty-factor
else
  return Optimization result
end if
```
The waiting costs for a spillback direction are multiplied by 'spillback-priority', leading to increased costs. For the experiments, the height of the penalty-factor is configured as 5, 10, or 1000. The increase in waiting costs has the effect that in the optimization, green time is allocated earlier to the spillback direction as it leads to the lowest overall waiting costs in the optimization horizon.

<span id="page-55-0"></span>

Figure 5.5: Case where spillback is detected, communicated, and processed by TFE (penalty X1000)

Example of a send/received spillback-state MQTT message from the Python module to the TFE traffic controller: {phase: 2, isSpillbackDetected: false,phase: 4, isSpillbackDetected: true, phase: 6, isSpillbackDetected: false}

# **5.4. Verification of implementation**

To verify the implementation of spillback detection. Singular experiments were executed to verify whether the queue length matches with the spillback detection [Figure 5.5,](#page-55-0) using the interactive mode of SUMO.When spillback is detected, a print statement is logged on the TFE side, including the calculated waiting costs. On the Python side, a print statement was sent to the terminal. The print statement included the phases, the queue length, the lane threshold, and whether spillback was detected as true or false.

**Example of logging**: Is Spillback detected: true. Cost of direction: 2144939.4129999997

The interactive mode of SUMO was also used to verify the override handling strategies and a penalty factor. In the same way, as with the spillback detection verification, spillback was enforced. After that, the override handling strategy was applied and analyzed whether the affected direction received green as soon as possible. When the penalty costs were incorporated, a penalty factor of times 1000 was used to analyze whether the waiting costs showed a significantly higher result than when a non-spillback queue was present at the intersection. In addition, an experiment was performed to validate whether more green time was allocated to a spillback directio[nTable 5.1\)](#page-55-1). The table shows that the amount of green time allocated increases when spillback occurs in the run on the west-side, and a spillback penalty is included.



<span id="page-55-1"></span>Table 5.1: Green time allocation with and without spillback penalty

In the case of the override strategy, the method first had trouble with allocating green and red to the correct locations, as the method allocated green to the affected direction without changing the traffic light in other directions. After this was addressed, the verification showed that when multiple directions were affected by spillback simultaneously, the override method had difficulties allocating green to the correct movement. The constraints of the override strategy were changed, so that only the spillback direction in combination with one other non-conflicting direction (instead of two) was allocated green time.

# <span id="page-55-2"></span>**5.5. Calibration of penalty factor in objective function**

As the simulation runtime per scenario is long, a decision has been made to either include the spillback penalty factor X5 or factor X10 in the corridor scenarios. This choice is made based on the total delay, and the spillback durations of the isolated intersection scenarios.

### **5.5.1. Calibration scenarios**

For the calibration of the penalty factor, experiments 1 to 12 (see [Table 5.2\)](#page-56-0) were used, and are similar to the first few experiments used to research effects at an isolated intersection, described in [chapter 6.](#page-58-0) First, in the base case no penalty factor was added when spillback is detected (experiments 1-3). Second, the spillback penalty factor was 5 (experiments (4-6), third, a penalty factor of 10 (7-9). Lastly, a penalty of 1000 (experiments 10-12), which resembles a maximum penalty when spillback is detected. The used simulation network for the calibration is similar to the intersection used for isolated intersection experiments. The configurations, regarding the simulation network and demand profile, are explained in more detail in [section 6.2](#page-59-0).

<b>Experiment</b>	<b>Demand</b> factor	Spillback control	<b>Note</b>			
		None	TFE.			
2	1.5	None	<b>TFE</b>			
3		None	Turning fraction (Turn) east 50/50			
4		Spillback penalty	Result X5			
5	1.5	Spillback penalty	Result X5			
6		Spillback penalty	Turn east 50/50 (X5)			
		Spillback penalty	Result X10			
8	1.5	Spillback penalty	Result X10			
9		Spillback penalty	Turn east 50/50 (X10)			
10		Spillback penalty	Result X1000			
11	$1.5\,$	Spillback penalty	Result X1000			

<span id="page-56-0"></span>Table 5.2: Overview calibration experiments

#### **5.5.2. Choice of penalty factor**

The results of average total delay, depicted in [Figure 5.6](#page-56-1), show five scenarios. The base case, the outcome of penalty factors X5, X10, X1000, and the control strategy override. The choice is made to include either the penalty factor X5 or the factor X10 into the corridor scenarios. Results showed an increase in total delay for all scenarios with a spillback control strategy, as depicted in [Figure 5.6.](#page-56-1) When the demand is increased with factor 1.5 the delay of the scenario with a factor of x5 increased more compared to the other scenarios. In the scenario with conflicting competing directions (see [Figure 5.6c](#page-56-1)) the overall delays do not show a significant difference between the spillback control strategy of a penalty factor of x10.

12 1 Spillback penalty Turn east 50/50 (X1000)

<span id="page-56-1"></span>

Figure 5.6: Overview average overall delays calibration scenarios

The spillback duration is improved when the penalty factor is X10 with demand factor 1.5, and spillback occurs (only) at

the west side of the intersection. When comparing spillback durations of the calibration scenarios (see[Table 5.3\)](#page-57-0), penalty factor X10 has lower spillback durations for the maximum affected direction (west), while disadvantaged direction do not have a significant decline, compared to penalty X5, except for the scenario with competing approaches. In further analyses, other modes are included, and the network size is increased to a corridor; the spillback control strategy is chosen as X10 as it showed promise on an isolated intersection and will have a more competing penalty with other types of road users in the corridor.



<span id="page-57-0"></span>Table 5.3: Spillback duration calibration results

# 6

# Experimental design

# <span id="page-58-0"></span>**6.1. Test environment**

The traffic system is translated into a software-in-the-loop simulation (SILS) to perform experiments. A SILS consists of a microscopic simulation, a virtual traffic controller running on the same computer, and a communication interface for information exchange between the microscopic simulation and the virtual controller [\(A. Stevanovic, Abdel-Rahim,](#page-85-10) [et al.](#page-85-10), [2009\)](#page-85-10). The virtual traffic controller runs the same software as the actual traffic controller. Information exchange between the microscopic simulation and the virtual controller is done by a communication module. The microscopic simulation is SUMO([Lopez et al.,](#page-83-7) [2018\)](#page-83-7), and the communication module retrieves traffic states, such as speed, position and traffic light states, using the SUMO Traffic Control Interface (TraCI). The retrieved data is then communicated to the traffic controller with an MQTT protocol. The communicated data is received and handled by the virtual traffic controller. After the traffic control optimization of this controller is complete, the proposed phase plan is communicated back to SUMO, which updates the traffic light states if necessary.

To verify the simulation runs, the total distance travelled for vehicles between the replications was compared to make a fair comparison between the alternatives([Table 6.1](#page-58-1)); these values were closely the same, as the simulation experiments is filled up for the first quarter (half demand), the demand was at peak level for the second and third quarter, and the last quarter the simulation was emptied again (half demand).

The utilized traffic system had random parameters; runs were compared with similar seeds to verify whether the results were the same. In the experiments, multiple random-seed numbers were used for vehicle arrivals, which departed according to a Poisson Arrival Process. Furthermore, ten replications per scenario were performed. For the prediction component in the utilized system, the results varied too much to analyze whether the difference in performance was due to a new random seed or the added spillback component. Because of the long simulation runtime, it was chosen to set



<span id="page-58-1"></span>Table 6.1: Total vehicle kilometers travelled in [km], base-case versus penalty times 10

a fixed random seed for the traffic controller component.

The experimental scenarios are distinguished by the used simulation network model, the included modes, demand, and the spillback handling method. The experiments are performed on two simulation networks. The first is an isolated intersection; the second network is a corridor consisting of ten intersections. In the following sections, the experimental design of these set-ups is described. Information about specific characteristics, such as measurements of link lengths and traffic control times, are reported in [Appendix B](#page-91-0).

# <span id="page-59-0"></span>**6.2. Isolated intersection**

#### **6.2.1. Simulation model**

The first test network is an isolated intersection with a left turn and straight direction for east and westbound traffic and two straight directions for north and southbound traffic. This way, effects such as lane-blocking are enforced in a more controlled environment and create a more isolated view of the effects of handling spillback effects modelling in traffic control. Further experiments also include a bus line configured with a stop 100 meters upstream of the intersection. The simulation model is depicted in [Figure 6.1a.](#page-59-1)

<span id="page-59-1"></span>

Figure 6.1: Simulation model and demand profile

### **6.2.2. Traffic light control and sensors**

In the current set-up, the traffic lights are constrained to the ring-and-barrier structure, as is explained in [section 3.3](#page-41-1). Because in this structure, the right and the straight direction are served green simultaneously, a left-turn bay is included in the network. Every traffic light is given a specific minimum and maximum green time, an extended green time of 2 seconds, a yellow time of 2 seconds, and a red clearance of 2 seconds. These specifications can be found in [subsection B.1.2](#page-91-1) in [Appendix B](#page-91-0). Lower minimum and maximum green time is allocated to the left turn due to lower demand than the straight directions. The same ring-and-barrier constraints apply to the adaptive control of the TFE. Consequently, four stage pairs can receive green at the same time: [2,6], [3,8], [4,8], and [4,7], the left turns cannot receive green at the same time in the corridor, that is why there is configured that they cannot receive green at the same time in the isolated intersection.

In the case of the TFE, additional traffic sensors are used. These sensors are constructed in SUMO by polygons([SUMO](#page-85-12), [2022](#page-85-12)) and can detect the exact vehicle position, speed, and length. Every lane has a specific polygon in the TFE. In the case of the isolated intersection, all polygons have the same lengths as the turning bays.

### **6.2.3. Modes and demand**

In the first several experiments the only road users are passenger cars, with varying demand. In further experiments, a bus service is added to see the effects of the optimization module on the traffic flow when multiple traffic modes are included. Furthermore, to see how the module copes with conflicting priorities in the optimization.

The bus service is set up as a bus that travels from north to south and halts for 20 seconds at the bus stop, 100 meters upstream of the intersection, indicated in [Figure 6.1a.](#page-59-1) Unless indicated otherwise, the turning fractions are determined at 75 percent for the straight directions and 25 percent for the left turns at the isolated intersection. The resulting demand profile for the scenarios is depicted in [Figure 6.1b](#page-59-1).

The demand profiles with factor 1 and 1.5 were chosen because spillback is enforced multiple times in the west side (on average 197 seconds for factor 1, and 174 for factor 1.5), while other direction still have tolerable queuing durations (5 seconds on average for north). The results showed large delays for the turning-bay on the east-side, it was suspected that the high delay was due to a low demand, consequently, a third demand profile was created in which the turning fraction of the east side was changed to 50% straight and 50% turn traffic, in addition, the demand for the was increased to 600 vehicles per hour.

# **6.2.4. Spillback handling**

Spillback occurs at least at one side of the intersection with the chosen demands. Every network configuration is tested without a spillback handling method, the override spillback method, and the penalty factor handling method. In the calibration in [section 5.5](#page-55-2), there was chosen to continue with a spillback priority of 10 for the 'minor' spillback penalty for the corridor.

# **6.3. Corridor**

# **6.3.1. Simulation model**

The corridor is divided into ten signalized intersections and has an arterial from north to south. The network is depicted in [Figure 6.2](#page-61-0) in which the corridor is cut in half, a full-size figure is depicted in [section B.2](#page-92-0). Intersection 3, Mission Bay, consists of a North-side and a South-side, named Mission Bay-North, and Mission Bay-South. Intersection 5, 3rd Campus, is located on the road's west side. Therefore, only traffic on the west side can travel over and turn right at this intersection. The LRV is indicated with a striped blue line, enters the corridor at intersection 1, and travels from north to south. The LRV leaves the network on the south side of the corridor. The LRV travels both ways.

### **6.3.2. Traffic light control and sensors**

[Figure 6.2](#page-61-0) indicates the directions in which cars can travel. Most directions follow the prescribed ring-barrier phasing, as indicated in [Figure 3.2a](#page-41-0) in [chapter 3.](#page-40-1) The phase times per intersection are depicted in the [Table B.4.](#page-95-0)

In the adaptive traffic light control of TFE, additional traffic sensors are used. In SUMO, these sensors are constructed with polygon zones and can retrieve the exact vehicle position, speed, and length inside the zone. Every lane has a specific polygon so that TFE can recognize which vehicle enters and leave a specific lane. In the same way, not only are LRV states retrieved on the dedicated lane but also pedestrians are identified at the corner of an intersection. Pedestrians can ask permission to cross the road by a push button.

# **6.3.3. Modes and demand**

In the corridor, cars, LRVs, and pedestrians enter and leave the network. Later scenarios simulate a mixed traffic flow by including trucks in the demand profile. Regarding the traffic demand, during the first quarter, the traffic participants enter the network with half the specified demand factor to fill up the network. In the second and third quarters, the demand factor is the specified demand, and in the last quarter, the demand factor is halved again.

For the corridor, the demand profile and the turning fractions per intersection([Table B.7](#page-99-0) and [Table B.6](#page-96-0) in [Appendix B\)](#page-91-0) are configured based on historical data, provided by Technolution.

<span id="page-61-0"></span>

Figure 6.2: Corridor San Francisco

# **6.4. Simulation experiments**

### **6.4.1. Route generation**

In SUMO, vehicle routes are generated based on demand definitions and junction turning percentages. In the router, a random-seed number determines the routes and the random departure times for flow input. The same seed numbers (1 to 10) are used between the replications for the departure profiles.

For the verification of the simulation runs, the total distance travelled for vehicles between the replications is compared. As the turning fractions cause a difference in routes per replication, the total distance travelled differs slightly in a more extensive network. However, this is not the case at an isolated intersection; in([Table 6.1](#page-58-1)), an example of the comparison is depicted.

# **6.4.2. Priority weights**

In the TFE algorithm, configurable parameters, such as the movement and object priority, influence the optimization, therefore, the outcome of the optimal signal plan. The object priority remains the same throughout the experiments and the chosen networks. For the corridor the standard the movement priority is indicated as 1 for every direction, in the corridor-priority scenario, additional priority was given to the corridor, as shown in [Table 6.2](#page-62-0).

# **6.4.3. Randomness**

In the model, two random components provide a different output every time. Therefore, a seed is specified for these parameters. The first is the predictive component of TFE, in which the model predicts whether the vehicle will go straight or turn in another direction based on the turning fraction. In the validation of the model, it was seen that the outcome was inconsistent with the same input. Therefore, the predictive component was configured with a fixed seed <span id="page-62-0"></span>Table 6.2: Object-priority and movement-priority configuration



(b) Movement-priority configuration



number.

The second random component is the arrival flow departures. Poisson arrivals are used between the different runs. The seed-number increases by one with each new replication and ten replications are made.

# **6.5. Network scenarios and experimental design**

Table 6.3: Generated network scenarios



During each experiment, spillback is enforced somewhere in the simulation model. For the isolated intersection, this results in the demand profiles in [Figure 6.1b](#page-59-1), for the corridor the demand profile is depicted in [Table B.7](#page-99-0). An overview of the experiments is written in the next section in [Table 6.4.](#page-63-0)

# **6.6. Analysis**

#### **6.6.1. Number of replications**

The number of runs depends on the uncertainty of the outcome; the number of replications should accomplish a statistical significance of 5 percent when comparing the means of a key performance indicator. However, a simulation run with MobiMaestro-Flow operates in real-time: 1 hour of simulation time is 1 hour in real-time. Therefore, ten replications are performed per scenario, leading to a runtime of ten to twelve hours per scenario.

Every run generates multiple output files by detectors placed in SUMO or SUMO generates the data itself. The keyperformance indicators travel time, delay, spillback time, and network performance are calculated from these output files. The method of data collection and generation depend on the simulation network.

#### **6.6.2. Additional sensors**

Instantaneous (E1) loop detectors are placed around the researched intersection for the detection of spillback duration upstream of turning bays([Figure 6.3](#page-64-0)). These sensors can detect vehicles standing still or travelling over the detector at <span id="page-63-0"></span>Table 6.4: Overview performed experiments



a specific speed. The detector gathers data from the travelling cars and assigns one of three states: *enter*, *stay*, or *leave*. The period of entering, staying, and leaving the E1 detector is measured. At direction where spillback could occur (sidestreets), E1 detectors are placed at an equal distance from the intersection, to calculate the queue durations at a certain distance for disadvantaged directions.

# **6.7. Key performance indicators (KPIs)**

The experiments collect data for three performance indicators at each replication. These three KPIs give insight into the effects of spillback modelling on the network's performance. The mean, standard deviation, and statistical significance



<span id="page-64-0"></span>Figure 6.3: Loop detector placed upstream of turning-bay

are measured for these indicators.

# **6.7.1. Spillback duration [s]**

The first performance indicator is the time spillback is experienced around the intersection, measured with an E1 detector in SUMO upstream of an intersection. When the vehicle queue flows over the detector with a speed lower than three m/s, the duration is calculated by the time the car enters until the car leaves the detector. The duration is measured in seconds. The total spillback time per direction is calculated as the sum of these durations.

At the isolated intersection, the detectors are placed just upstream of the turning bays on the east and west side of the intersection. The detectors are placed at an equivalent distance for the north and south of the intersection. Important to note that spillback duration is measured in the north and south directions; however, it is used for measurements and in every scenario, no spillback control strategy is applied in the north and south directions.

In the corridor, E1-detectors are placed upstream of the northernmost intersection 3rd-Channel; furthermore, the detectors are placed at the north, south, east and west sides of the intersection 3rd and 16th Street. In the same way as the isolated intersection, just upstream of the turning bay or at the directions with no turning bay: at a similar distance.

# **6.7.2. Delay**

The second performance indicator is vehicle delay in seconds, which is a derivative of the travel time. The travel time is measured using vehicle route duration. At the isolated intersection, this leads to six routes: vehicles travelling from **(1)** east to west, **(2)** west to east, vehicles travelling from **(3)** east to south or **(4)** west to north, and vehicles travelling from **(5)** north to south or **(6)** south to north. For the corridor, six routes are selected that are affected by spillback or by the spillback control strategy, indicated in [Figure 6.4.](#page-65-0) Routes 1-4 are for passenger cars; routes 5 and 6 are the LRVs travelling from south to north or from south to north, respectively.

The individual delay is measured as the vehicle's travel time to complete its route minus the travel time of the route when the vehicle is travelling with free-flow speed. As the free-flow speed travel time is a fixed variable per route, the data distribution of the travel time and the delays are similar. The free-flow travel time is calculated by the route length divided by the free-flow speed (13.41 m/s).

$$
Vehicle delay = Traveltime - Free-flow\,
$$
\n
$$
(6.1)
$$

### **6.7.3. Overall intersection delay**

For the network indicator, the overall intersection delay is calculated. The method differs between the isolated intersection and the corridor. For the isolated intersection, the overall delay is calculated as the sum of the direction delay.

<span id="page-64-1"></span>Calculating the total delay in the same way as the isolated intersection is rather complex, as it requires multiple additional sensors per intersection and a lot of data processing. Therefore, the total delay is calculated differently. The delay per intersection is estimated by SUMO, defined by the sum of the time loss and the waiting time. The waiting time is the time spent waiting for a vehicle. The time loss is the time a vehicle travels below the ideal speed during their trip. The disadvantage of these variables is that a vehicle stop at traffic lights is defined as a scheduled stop and is not considered in this delay calculation.

<span id="page-65-0"></span>

Figure 6.4: Routes compared in the analysis

#### **6.7.4. Statistical significance**

The statistical significance is calculated by a variation of the student t-test, namely the Welch t-test. Welch's test is regarded as the parametric equivalent of the Two-Sample t-test. This test is chosen, because the variances of the compared means are assumed not to be equal. This is compensated by a different method of approximating the degrees of freedom. The means of the two samples are compared and tested to determine whether a hypothesis is accepted or rejected. The following hypothesis is tested:

 $H_0$ : There is no difference between the means of scenario 1 (basecase) and scenario 2 (comparison-scenario).

*H*<sup>1</sup> : The means of scenarios 1 (base case) and 2 (comparison-scenario) are different.

First, the means  $\overline{x}_1$  and  $\overline{x}_2$ , and the variance  $s_1$  and  $s_2$  are calculated. After that, the difference of the variance between the means is calculated using the number of replications  $n_1$  and  $n_2$  as follows:

$$
s_{\text{diff}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}
$$
 (6.2)

The *t*-value is then calculated as the difference in the scenario means, divided by  $s_{\text{diff}}$ :

$$
t = \frac{\overline{x}_1 - \overline{x}_2}{s_{\text{diff}}} \tag{6.3}
$$

After that, the degrees of freedom *df* are calculated, using the following formula:

$$
df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{s_1^2}{n_1}\right)^2 + \left(\frac{s_2^2}{n_2}\right)^2}
$$
(6.4)

The degrees of freedom lead to a t-value from the t-table. The null hypothesis is rejected if the calculated t exceeds the t-value from the t-table.

A two-tailed significance of  $\alpha = 0.05$  is chosen, and the number of replications per scenario is 10. The following example table compares the base-case scenario with the scenario where a spillback penalty factor (x10) is included. When the significance of  $\alpha = 0.05$  is not reached, and the result is  $p < 0.1$  or  $p < 0.25$  it is mentioned in the result tables. When the *p*-value is above 0.25, it is declared insignificant, meaning no difference is found between the means of the experiment compared to the base case.

Table 6.5: Outcome delays base-case (experiment 5)



Table 6.6: Overview KPIs experiment 11



# Results

 $\overline{\phantom{a}}$ 

The experimental scenarios led to quantitative results that indicate the effect on the traffic and network performance when spillback control strategies were incorporated in traffic light control optimization. A Welch t-test was used, as explained in [subsection 6.7.4,](#page-64-1) to test the statistical significance of the hypothesis if there is a difference in means between the scenario without a spillback control strategy (base case) versus cases where a spillback control strategy was implemented. When a result is significant, the p-value results in a lower value than 5 percent. In the following sections, key results are described, first on the effects on the network performance at an isolated intersection and, after that, on the network performance in a corridor. An overview of the performed experiments is found in the previous chapter, in [Table 6.4](#page-63-0). The generated graphs and tables can be found in [Appendix C](#page-100-0) for unimodal experiments at an isolated intersection, in [Appendix D](#page-114-0) for isolated intersection scenarios with a busline, and in [Appendix E](#page-124-0) for the results of the corridor scenarios.

# **7.1. Isolated intersection**

The experiments performed at an isolated intersection (see [Table 6.4\)](#page-63-0) varied in the type of road users, the demand, and the spillback control strategy. In the next sections, the results of these experiments are described, first, the overall intersection delay is described, and then the spillback duration and delay performance. After describing the results of different scenarios, the results are discussed.

# **7.1.1. Passenger car experiments**

For all scenarios, the overall intersection performance decreased when a spillback control strategy was applied. In the experiments with the turn ratio set to 50/50, the total intersection delay results were less spread, however, the overall delays still increased. The highest intersection delay in the scenarios was found for the strategy override.

#### Spillback duration

In general, the override strategy showed the most significant reductions for all directions affected by spillback (*p* <0.05). However, in most cases, this was paired with an increase in delays for the directions that were not bound by spillback. For example, in the scenario turn ratio 50/50, increases showed increases of 146% and 86% in queue durations for the north and south.

With a demand factor of 1, all spillback penalty strategies showed a decrease in spillback duration for the west, with the most significant decrease for the penalty factor X10. The total spillback duration also decreased for factor X10, which is also the case for the larger factor x1000; however, the data of the replications are more spread out.

When the demand was increased with factor 1.5, the penalty strategies did not show a statistical difference compared to the base case; however, the data in [Table 7.1](#page-68-0) shows decreased spillback durations for the east side. The direction with the highest demand, the west, increased in spillback duration.

When the east-side demand was increased, and also turn ratio was altered to 50/50, improvements were found for the





(a) Experiments with increased demand x1.5 (b) Experiments east turning fraction 50/50 and higher demand

Figure 7.1: Overall mean average delay

<span id="page-68-0"></span>Table 7.1: Results demand factor 1.5 and turnratio 50/50



west side for all spillback control strategies (*p* < 0.05). However, this was paired with increased spillback durations for the east side. With a high penalty factor, the north side benefits with a reduced queue duration compared to the base case. However, duration at the south increased ( $p < 0.05$ ).

#### Delays

With low demand, the delay decreased with all spillback strategies and demand scenarios, for the direction in which spillback was enforced (west). With an increased demand factor of 1.5, the delays of the directions with a low traffic demand increased significantly (turning-bay east side, north and south side). In contrast to other directions, as with a minor penalty factor, the delays decreased for north, east, and west, while the turning-bay delay stayed similar.

In the scenario with a turning ratio of 50/50 on the east, a larger penalty factor shows more improvements for directions that are affected by spillback, but this is paired with increased delays for the north and the south. However, with a very high penalty factor (X1000), the turning bay of the east side is skipped more often.

# **7.1.2. Experiments with busline**

In the experiments with low demand, the overall intersection delay was decreased when spillback penalty strategies were applied ( $p$ <0.05), as shown in [Figure 7.2a.](#page-69-0) The most significant decrease is with the highest penalty factor of X1000. The decrease in overall intersection delay disappeared (see [Figure 7.2b](#page-69-0)when the demand was increased with factor 1.5 (*p* < 0.05), the means of the spillback penalty strategies were comparable. For both demand profiles, the override strategy increased the overall intersection delay.

<span id="page-69-0"></span>

Figure 7.2: Total mean delays from bus experiments

#### Spillback duration

In terms of spillback durations, the most considerable reductions are found with the override control strategy, as shown in [Table 7.2.](#page-70-0) But this is also paired with the highest increase in queue durations for disadvantaged directions north and south.

With a demand factor of 1, the penalty strategies reduced spillback durations. The most significant decrease is found for the penalty factor X10. The benefit is decreased when the demand factor is increased to 1.5; however, spillback is still reduced for the west side. For other directions, the queue or spillback duration is increased.

#### Delays

Regarding the delays, reductions were found for the straight and turning bay of the west side with low demand and a penalty factor of X5 (see [Table 7.2\)](#page-70-0). Additionally, minor reductions for the north, south, and east side. Also a reduction was found for the bus delay. In general, with low demand, the delay of the bus was decreased for all spillback penalty strategies.

The benefit decreased when the demand was increased with factor 1.5. Then, the reductions in spillback duration were found to be the highest for the spillback penalty X10. No benefit is found for a smaller penalty factor of X5, only a slight decrease for the west side. Furthermore, the disadvantaged directions, the north and the south, have no significant decrease in performance, except for the east side, which has increased delays when the demand is higher, and also the delays of the bus increased with a spillback control strategy applied.

# **7.2. Interpretation results – isolated intersection**

With a low demand, the overall intersection delay increases when spillback penalty factors are included in the traffic light control optimization. In a scenario where one direction has an apparent higher demand, the provided additional priority does does not compensate the loss in delays for other, disadvantaged, directions, leading to a lower intersection

		Mean base case		Mean $(X5)$		Mean (X10		<b>Mean (X1000)</b>		Mean (Override)	
		1	1.5	1	1.5	1	1.5	1	1.5	1	1.5
Spillback	North	10.188	11.697	8.6%	$-20.0%$	10.8%	6.5%	$-0.3%$	19.3%	79.3%	64.9%
	South	6.846	7.982	5.8%	$-2.7%$	24.9%	13.2%	30.6%	11.8%	133.7%	121.7%
	East	13.381	14.735	72.5%	23.7%	30.8%	$-0.7%$	29.3%	32.5%	$-41.3%$	$-25.3%$
	West	205.849	227.57	$-15.9%$	$-8.6%$	$-12.1%$	$-17.1%$	$-2.3%$	$-7.0%$	$-50.4%$	$-56.4%$
	Total	236.264	261.984	$-9.2%$	$-7.2%$	$-7.7%$	$-14.4%$	0.5%	$-3.1%$	$-39.4%$	$-44.6%$
Delays	North	31.921	31.844	$-0.9%$	3.4%	0.2%	2.3%	1.7%	4.1%	7.7%	9.5%
	South	34.701	34.776	$-0.3%$	1.2%	0.2%	2.2%	3.4%	3.7%	14.3%	12.0%
	East	25.125	23.186	$-1.1%$	1.8%	$-0.7%$	7.9%	$-5.0%$	3.6%	$-0.9%$	16.4%
	West	22.523	22.933	$-5.1%$	$-1.5%$	$-4.3%$	$-4.5%$	$-0.4%$	$-1.4%$	$-14.4%$	$-17.8%$
	TB: East	68.376	64.018	2.8%	3.8%	$-1.5%$	3.8%	$-6.1%$	4.9%	1.3%	2.8%
	TB: West	54.551	54.93	$-2.1%$	2.6%	$-5.9%$	$-4.5%$	$-4.6%$	$-2.3%$	$-4.4%$	$-2.0%$
	Bus	37.9	28.72	$-17.9%$	8.7%	$-7.6%$	26.4%	$-23.7%$	9.5%	$-5.1%$	25.6%

<span id="page-70-0"></span>Table 7.2: Result bus experiments isolated intersection

performance. Furthermore, when a spillback penalty strategy is applied, the direction affected by spillback, receives green to minimize the waiting costs, the non-conflicting direction that is allocated green time is the direction with the highest demand, receiving priority over other non-conflicting directions. This caused high delays for directions with a low demand sharing a barrier with a direction where spillback occurred. For example, the turning-bay on the east side, with a low demand. chosen

With a high spillback penalty, a direction with spillback receives green as fast as possible; however, the penalty is already given when the traffic exceeds the turning bay; this way, absolute priority is given very fast, and the traffic light switches more often, while other direction also have logner queues. Switching the traffic light takes time due to providing sufficient clearance, which can lead to an increased spillback duration (10X versus X1000).

The override strategy showed promising results on spillback duration; even with approaches competing in demand, benefitss were shown for both the east and the west side. This is caused by the abrupt method. However, the method does not consider safety constraints, such as red clearance. This results in a dynamic phasing situation, switching traffic lights when spillback is detected as true. After that, the traffic light control works as usual. The absolute prioity also led to very high delays for disadvantaged directions.

When a busline is is included in the network, delays decrease for all directions when the traffic demand is low and one side is enforced with spillback. This caused the traffic lights to switch more often: when spillback is detected, the west side received green, and when a bus is detected at the north-side green is allocated faster as the object priority of a bus is twenty times higher. This caused a lower bus delay and lower delays for spillback directions. However, the benefits are lost when a conflict occurs in traffic control. For example, with a rising traffic demand: no difference with the base case or worse performance was found in delays and spillback durations. With a small penalty factor, improvements are still found for spillback and north and south directions, as the high object priority of the bus is more in balance with a queue of vehicles that overflows the turning bay. Though, the benefits are paired with an increase in bus delay. With increased penalty factors, traffic control leans into an advantage for spillback directions and larger delays for the bus (and non-spillback directions).

# **7.3. Corridor**

In experiments 26 to 37, experiments were performed on a corridor. The experiments have the same demand profile, but vary in priority given to the corridor, the spillback control strategy, and the included traffic on the road. First, results are described for the scenario with cars, LRVs and pedestrians. Next, the corridor is assigned 10 percent additional movement priority. After that, the result are presented for the last scenario, when the traffic flow is more mixed with the inclusion of buses. First, the intersection performance and network performance are described. After that, the spillback duration is described for the intersection 3rd and 16th Street, which is measured by detectors placed around the intersection (see [Table 7.3b\)](#page-71-0). Furthermore, the delays for the LRV and different vehicle routes are described (see [Table 7.3a\)](#page-71-0). An overview of the intersections on the corridor is shown in [Table 7.4.](#page-71-1)



<span id="page-71-0"></span>(a) Routes for delay analysis

Table 7.3: Overview routes and detector locations

<span id="page-71-1"></span>Table 7.4: Intersections in the corridor



# **7.3.1. Traffic flow with car, LRV, and pedestrians**

When spillback control strategies were applied, the network performance, which is the sum of all intersection performances, depicted in [Table 7.5](#page-71-2), showed increased performance when the spillback penalty was configured as X10. With a very high spillback penalty (X1000), the network performance decreased. The highest performance increase is with the control strategy override, however, this strategy neglects the ring-and-barrier structure when controlling the spillback effects.



<span id="page-71-2"></span>Table 7.5: Network performance with normal demand in delay [s] and [%]

When looking at the individual intersection performance (see [Table 7.6\)](#page-72-0), intersections 6. 3rd 16th Street and 7. 3rd Mariposa have the highest delays compared to the other intersections. With a penalty strategy applied, benefits are found for 3rdMariposa, which is the intersection directly south of 3rd and 16th Street. Furthermore, at the ends of the corridor, benefits were found, mainly during the peak period Q2/Q3. At the intersection where spillback was enforced, 3rd and 16th Street, the overall intersection delays increased when spillback control strategies were applied.

Outside the peak period (Q1 and Q4), a decreased performance was found for several intersections when a higher spillback penalty of X1000 was applied. These delays were higher compared to a lower spillback penalty – for example, in Q1, the delay of intersection 4. 3rd Warriors Way, increased 20.9% compared to 0.6 % when a spillback penalty of X10 was applied.

Regarding the spillback effects at 3rd and 16th([Table 7.7](#page-72-1)). With spillback penalty strategies were applied, the spillback duration of the north decreased, and decreased more with a higher spillback penalty. In contrast to the overflow, occurring upstream north of 3rd and 16th, is handled faster when no spillback penalty strategy was applied. With the current demand profile, there were increased queue durations for disadvantaged directions, however, these durations are not that long.

Regarding the route delays, routes travelling from north to south, such as routes 1 and 4 decreased when the penalty factor was X10 ( $p$ <0.05). These routes are the routes that benefited from the spillback control strategy. Route 2 also had a decreased delay and travelled from south to north. Only route 3, which travels from the east side of 3rd and 16th
Table 7.6: Intersection performance in delay [s] and [%]



street, has an increased delay. The delay increased when the spillback penalty was high (X1000); furthermore, the benefit for route 4, travelling from the west of 3rd and 16th to the south of the corridor, is lost with a higher spillback penalty. The delays for the LRV are not statistically different from the basecase, however, the LRV travelling to north has a small decrease in delay for all cases where a spillback control strategy is applied.

Table 7.7: Results of experiments with normal demand

		<b>Basecase</b>	<b>X10</b>	X1000	<b>Override</b>	
	KPI	Mean [s]	Mean [s]	Mean [s]	Mean [s]	
Spillback	Overflow	7.208	118.9%	37.2%	$-100.0%$	
	North	212.99	$-29.5%$	$-35.5%$	$-84.1%$	
	South	2.363	29.4%	11.1%	40.6%	
	East	3.143	33.1%	8.9%	78.5%	
	West	5.981	$-17.7%$	$-0.5%$	$-16.6%$	
	Total	224.477	$-27.7%$	$-33.5%$	$-78.7%$	
Delay	Route 1	197.316	$-6.3%$	$-9.0%$	$-21.0%$	
	Route 2	175.867	$-1.8%$	$-3.3%$	$-8.8%$	
	Route 3	115.71	4.7%	12.5%	9.7%	
	Route 4	116.986	$-4.5%$	1.5%	$-7.2%$	
	LRV-north	225.823	0.2%	$-1.3%$	$-0.7%$	
	LRV-south	207.86	0.5%	2.7%	2.9%	

### **7.3.2. Corridor priority**

When corridor priority is configured for the arterial, vehicles leaving, entering or travelling over the arterial receive an increased priority. The corridor in combination with a spillback penalty strategy reduced the network performance. For both spillback penalties X10 and X1000, the overall network delay increased, furthermore, a higher network delay was found when the spillback penalty was higher. However, applying the override strategy showed a slight improvement on the network performance.

Table 7.8: Network performance corridor priority in delay [s] and [%]

	<b>Basecase</b>	<b>X10</b>	X1000	Override
01	328.6	$-0.84%$	3.04%	$-1.57%$
02/03	664.1	0.45%	2.01%	1.03%
04	346.9	0.63%	6.51%	$-2.28%$
Total	1339.6	በ 18%	3.43%	$-0.46%$

Regarding intersection performance, (see [Table 7.9\)](#page-73-0) the outer intersections still had an increased performance when spillback penalty strategies were applied. With a higher spillback penalty, the performance benefits were mainly found during the peak-demand in Q2, in comparison to Q4 which had decreased performances for all intersections. The increased corridor priority for the arterial, in combination with the additional spillback priority, did not lead to better performances than without the corridor priority.

	<b>Base case</b>				Result X10			Result X1000		<b>Result Override</b>		
<b>Intersections</b>	Mean 01	<b>Mean 02/03</b>	Mean 04	Mean 01	<b>Mean 02/03</b>	Mean 04	Mean 01	<b>Mean 02/03</b>	Mean 04	Mean 01	<b>Mean 02/03</b>	Mean 04
0. 3rd channel	24.765	49.494	24.599	$-4.1%$	$-2.7%$	3.2%	$-10.4%$	$-5.8%$	5.1%	$-9.9%$	$-11.6%$	$-2.5%$
1.3rd missionrock	14.742	34.812	18.659	22.7%	8.6%	$-2.3%$	5.7%	9.5%	3.7%	5.6%	4.3%	$-12.7%$
2. 3rd_missionbaynorth	15.027	25.995	15.39	$-9.6%$	2.6%	$-5.0%$	$-7.6%$	$-1.8%$	0.1%	$-10.0%$	1.5%	$-3.1%$
3. 3rd missionbaysouth	22.995	38.899	20.897	$-17.0%$	4.5%	1.8%	$-5.9%$	1.9%	7.2%	$-2.4%$	4.7%	$-1.4%$
4. 3rd warriorsway	19.501	37.382	16.209	$-16.2%$	6.8%	$-1.5%$	$-2.6%$	8.6%	9.6%	$-2.3%$	2.3%	7.6%
5. 3rd_campus	33.22	66.779	32.973	8.8%	3.4%	$-1.6%$	9.4%	4.0%	7.8%	7.4%	$-1.0%$	5.1%
6. 3rd 16th	52.826	111.584	57.659	1.6%	1.7%	1.8%	1.4%	7.8%	9.3%	$-8.3%$	7.1%	$-0.9%$
7. 3rd_mariposa	54.89	120.981	68.442	1.8%	$-0.6%$	6.5%	7.7%	$-1.7%$	2.8%	1.0%	$-8.4%$	$-11.0%$
8. 3rd 18th	36.554	76.455	37.714	$-2.5%$	$-6.7%$	0.0%	7.5%	$-0.6%$	8.6%	5.1%	2.4%	$-0.5%$
9. 3rd_19th	26.757	47.603	23.746	$-0.9%$	$-1.8%$	0.4%	0.3%	4.0%	16.4%	$-4.4%$	15.3%	5.0%
10. 3rd 20th	27.356	54.079	30.641	$-0.4%$	$-2.3%$	$-8.4%$	13.8%	$-2.5%$	1.8%	$-1.3%$	3.3%	0.2%

<span id="page-73-0"></span>Table 7.9: Intersection performance in delay [s] and [%]

In general, spillback durations were already reduced due to the increased corridor priority, at the north of 3rd and 16th Street. Nevertheless, the penalty factors have decreased the spillback durations even more for the north and south (see [Table 7.10,](#page-73-1) also shorter queue durations were found in the east direction. The decrease in total spillback, had the effect that the west side of 3rd and 16th Street had a very high increase in queue duration. The duration increased even more when a higher spillback penalty was applied. Furthermore, for the override strategy, the biggest reduction in spillback duration, was paired with the highest increase in queue durations for the south, the east, and the west.

<span id="page-73-1"></span>Table 7.10: Performance indicators with corridor priority



Regarding the route delays, all routes have received additional priority as they were travelling over the arterial. With a penalty factor of X10, routes 2 and 3, travelling to the north of the corridor had decreased delays, but also vehicle travelling from the east of 3rd and 16th Street to the south of the corridor (route 4). With a higher penalty factor (X1000), only a decrease in delay was found for the third route. Regarding the LRV delays, no high difference in delay was found. The LRVs travelling to the south had decreased delays when spillback penalties were applied, which was not the case when no corridor priority was applied in the previous scenario.

### **7.3.3. Mixed traffic flow with bus**

For these scenarios with a mixed traffic flow, buses are included and travel on the same road as cars. A bus receives more priority than a normal passenger car, consequently, when a bus is in a queue for the trffic light, the queue has extra priority, compared to a queue consisting only of cars. Regarding the network performance (see[Table 7.11\)](#page-74-0), the spillback control strategies caused an increased network performance during the peak demand (Q2/Q3). With a higher spillback

penalty, the benefit was also seen for the period after the peak in Q4. For the total network delay, there was no difference, or a small improvement in network performance for higher spillback penalties.



<span id="page-74-0"></span>Table 7.11: Network performance mixed traffic flow in delay [s] and [%]

Regarding the intersection performances, shown in [Table 7.12](#page-74-1), the increased performances are more found at intersections in the middle of the corridor, for example 7. 3rd Mariposa, and 3rd Warrior's Way, instead of outer intersections, such as 0. 3rd Channel, 1. 3rd Mission Rock, and 3rd and 9. 19th Street, which had increased performances in previous scenarios. However, when the spillback control strategy was override, the absolute priority did increase the network performance of the four northernmost intersections (*p*<0.05) paired with southernmost intersections 7. 3rd Mariposa, 8. 3rd 18th and 10. 3rd 20th (*p* < 0.05).

<span id="page-74-1"></span>Table 7.12: Intersection performance with mixed vehicle flow, delay in [s] and [%]

	<b>Base case</b>			Result X10			Result X1000			<b>Result Override</b>		
Intersections	Mean 01	<b>Mean 02/03</b>	Mean 04	Mean Q1	Mean Q2/Q3	Mean 04	Mean Q1	Mean Q2/Q3	Mean 04	Mean 01	<b>Mean 02/03</b>	Mean 04
0. 3rd channel	26.174	54.309	29.184	$-2.2%$	0.5%	8.0%	1.3%	$-1.9%$	2.8%	3.2%	$-12.3%$	$-3.3%$
1. 3rd missionrock	17.187	39.648	21.864	13.1%	6.2%	0.2%	12.0%	10.1%	$-7.0%$	1.7%	$-5.5%$	$-4.9%$
2. 3rd_missionbaynorth	14.953	31.833	15.743	$-2.0%$	$-3.9%$	10.9%	0.4%	$-4.4%$	$-4.3%$	$-1.9%$	$-6.9%$	5.3%
3. 3rd missionbaysouth	22.401	50.721	24.777	$-3.0%$	$-9.1%$	2.9%	0.4%	$-9.9%$	$-1.3%$	0.9%	$-4.1%$	0.6%
4. 3rd_warriorsway	22.798	41.354	19.998	$-11.4%$	5.8%	10.6%	$-8.1%$	2.3%	$-9.6%$	$-8.4%$	5.6%	$-9.9%$
5. 3rd campus	35.227	78.418	37.686	11.3%	$-4.5%$	$-2.2%$	$-1.4%$	$-7.0%$	$-1.4%$	4.5%	$-4.3%$	4.7%
6. 3rd_16th	55.782	136.641	62.552	$-0.9%$	$-2.5%$	4.4%	1.1%	1.4%	2.5%	11.4%	0.3%	4.1%
7. 3rd_mariposa	62.859	137.685	76.758	14.7%	$-1.9%$	$-7.4%$	0.8%	$-5.4%$	$-6.4%$	1.0%	$-10.2%$	$-8.0%$
8. 3rd 18th	42.999	86.442	43.639	13.0%	$-5.0%$	$-4.3%$	$-1.3%$	$-1.9%$	$-5.3%$	8.2%	$-4.4%$	$-9.7%$
9. 3rd 19th	33.332	53.355	27.88	$-13.7%$	1.3%	5.8%	$-3.3%$	6.6%	$-3.3%$	$-3.5%$	7.7%	$-2.5%$
10. 3rd_20th	29.866	64.279	33.562	1.9%	$-1.6%$	9.2%	1.0%	0.0%	4.4%	1.6%	$-4.8%$	11.4%

In terms of spillback duration, shown at the top of [Table 7.13,](#page-74-2) applying a spillback control strategy lead to high reductions in spillback duration on the side in which spillback was enforced (north-side of 3rd and 16th). But also, a high reduction was found for the west side, when a small spillback penalty of X10 was applied. Furthermore, a higher spillback penalty (X1000), did not lead to higher reduction, but caused longer queue durations for disadvantaged directions.

<span id="page-74-2"></span>Table 7.13: Results of corridor experiments with bus



The delays of the routes are shown at the bottom of[Table 7.13](#page-74-2). The effect of spillback control strategies was negative for route 3, which is the route from the east side of 3rd and 16th Street to the north of the corridor. In contrast to vehicles travelling from north of the corridor to the south, or 3rd and 16th Streets to the south (routes 1 and 4). Regarding the LRV, there was little to no difference in delay for the line travelling to the north. The LRV travelling to the south did had increased delays for all spillback control strategies.

### **7.4. Interpretation results – Corridor**

Every scenario had the same demand profile, in which spillback was enforced on the north side of 3rd and 16th Street and also overflow occurred in some instants at 3rd Campus, just north of 3rd and 16th Street. The results showed with a moderate vehicle flow on the road and a minor spillback penalty of X10, the overall network performance was increased. Furthermore, during the peak period (Q2/Q3), the network performance improved for all scenarios with spillback control strategies, except for the scenario with additional corridor priority. Furthermore, improved network performance was found when absolute priority was provided and the suggested phase plan was interrupted with the override strategy. Regarding intersection performance, intersections also benefit the most from the override strategy, provided that the traffic demand gradually reduces. Although, these promising results are due to the abrupt spillback control by neglecting phase constraints, in contrast to the 'TFE' spillback penalty control strategy.

Regarding the intersection performance, intersections with lower demand or fewer conflicting directions gain performance when spillback penalty strategies are applied. For example, 3rd Warriors Way or 3rd Campus, which are both three-legged intersections. Intersection decreased in performance when multiple directions had a higher demand. It can be said that the overall intersection delay decreases when the demand for conflicting directions is low, consequently, when the demand for conflicting directions is high, the gain in overall intersection performance is lost. Still, the spillback durations are reduced as these movements receive priority over other directions.

At the investigated intersection (3rd and 16th Street), spillback was reduced more with a stronger spillback control strategy. This was paired with increased delays for non-spillback directions. However, when spillback occurred more often on the north side, the control allocated green time more often to the north side, but the queue duration on the west side also decreased. The west side is in a separate phase barrier, and due to a higher demand the west receives priority over the other direction in the same barrier, the east side, consequently, the phases of the west side are always the successive phase of the spillback direction (north). Furthermore, the performance of the intersection south of 3rd and 16th improved when spillback control strategies were applied. The intersection downstream of 3rd and 16th has a long approach link, which meant that the vehicle platoons approaching 3rd Mariposa often received a green wave, compared to when no spillback control strategy was applied. This could be due to an increase in vehicle platoons, or in the case of no green wave, the lanes in front of the intersection are immediately detected with spillback when the vehicles arrived (overflow of turning-bay). Consequently, this direction receives high priority, compared to the other directions at the intersection.

Regarding the overflow measured upstream of 3rd and 16th Street, a small spillback penalty does not provide the proper priority to handle the spilled-over vehicles. This is because of a short link where the overflow occurs on. Consequently, in the scenarios it could be seen that only a larger spillback penalty (X1000) or providing absolute priority did reduce the overflow duration.

When interpreting the results of combining a spillback control strategy with a movement priority for the arterial (corridor priority), the 'extra' priority did not lead the more benefits when only a spillback control strategy was applied, as the additional priority led to higher delays for disadvantaged directions and the reduce in spillback duration was not significantly different compared without the corridor-priority.

In terms of public transit, the impact of spillback control strategies on the delay of Light Rail vehicles depend on the weight of the spillback penalty: passenger cars entering the network at 3rd Channel (with spillback priority) can be given the right of way when the spillback penalty is high, compared to the LRV with a significant higher object priority. A high spillback penalty or the override strategy leads to higher LRV delays. The average delay for the LRV is still manageable, as the highest LRV delay for the scenarios without a bus is 3.1% with a high spillback penalty (X1000). However, mainly the delay occurred at the north side of the corridor, as this was the only conflicting direction, in most cases the LRV benefited (or no difference), because same way as arterial. The same

When the traffic flow was mixed with a bus, low additional delays and spillback durations are found, as the bus promotes the traffic flow of the North with the additional vehicle priority. This was paired with higher delays for disadvantaged directions. As the bus shares the same infrastructure as the vehicles, effects are not that different from scenarios with only cars. However, it should be noted, that when a bus travels over a conflicting directions, for example from east of 3rd 16th Street to the west of 3rd 16th Street, this lead to conflicting control objectives due to a comparable waiting cost.

Consequently, these situations leads to a larger intersection delay.

The results at the corridor showed that a penalty factor effectively decreased the spillback durations, with a minor increase in delay for the public transit. When the spillback penalty was minor (X10), the cost for disadvantaged directions was a small increase in average delay. However, it should be mentioned that this is the effect when other directions have a relatively low demand, compared to the spillback direction. When multiple directions have high demand, the benefit is lost. Furthermore, when the demand was factor 1 during Q2/Q3, the performance of the whole network increased, but in half demand (Q1 and Q4), in most cases, the performance decreased a bit. The results also showed that having a fixed penalty factor throughout varying demands, the penalty factor may not always be the right weight. For example, overflow occurring at a short link does not receive the proper priority when a minor spillback penalty is applied. However, a minor spillback factor to add additional priority to spillback, but no absolute priority should be given.

In the next chapter, the results of the research are concluded, and the research questions from this thesis are answered. Furthermore, the limitations surrounding this research are discussed.

# 8

# Conclusions and recommendations

This chapter answers the questions posed at the beginning of the thesis. It was found that little research was done on spillback control in multimodal RTCS system and knowledge was lacking on the effects of spillback control in a RTCS with multimodal optimization. This research focuses on gaining insight on the effects of spillback control in multimodal RTCS so that gap is filled. Therefore, the first question (RQ1) was:

1. *How can an existing RTCS be extended with spillback detection and spillback control strategies in a multimodal context?*

To research the effects of the spillback control strategy, simulation study was constructed, with the central research question being as follows:

2. *What is the effect of controlling spillback in a multimodal RTCS at signalized intersections on the traffic and network performance?*

The main research questions were divided into multiple subquestions. The following sections conclude the outcome of the literature review and the design choices for the component, followed by a discussion. After that, the effects of the spillback control strategy on an isolated intersection and a corridor are concluded and discussed (RQ2).

# **8.1. Research question 1 – Literature research**

Research question 1 was divided into four subquestions. Namely how spillback could be detected (RQ1.1), how spillback could be controlled (RQ1.2), and the identification of the architecture of RTCS with multimodal optimization (RQ1.3). After that, there was explored on how a spillback detection method and control strategy could be implemented in a multimodal RTCS. The findings are concluded in the next sections.

### **8.1.1. Input for spillback component design**

### Spillback detection

In the literature was found that the spillback detection method depended on the utilized traffic sensors. The most pragmatic approachwas using video sensors that supervisewhether spillback occurs at an inner junction, or lane blockage at a turning bay, which performs well in good weather conditions. When sensors were used that can count vehicle arrivals and speeds (e.g. loop detectors), spillback was detected in the form of a threshold. In these approaches, a queue method was used to estimate whether the capacity was exceeded. More specifically, shock wave theory was applied to estimate queue lengths in real-time.

#### Spillback control strategies

The research papers described mainly two approaches to control spillback. The first approach was a form of metering by limiting the inflow with a reduction in green time for critical phases at upstream intersections, the green time extension

for the congested link, or a combination. The second approach controls spillback with an optimization algorithm to minimize an objective function that avoids or handles spillback queues. Possible criteria are minimizing the queue length, maximizing the throughput, minimizing the delay, or finding an overall optimal flow for all approaches.

#### Multimodal RTCS architecture

A real-time controlled traffic system with multimodal optimization optimizes multiple vehicle flows based on an objective criterion. Traffic sensors measure traffic input, such as vehicle position and speed, for multiple modes. The sensor input is processed and used for modelling. The modelling part of the RTCS consists of estimation and prediction of traffic variables. For the estimation or prediction of traffic variables, the delay is used to optimize for multiple road users, which can be either personal delay or vehicle delay. For passenger cars, the delay is often calculated according to a inputoutput method with a vertical queuing model, but also systems were identified that apply shockwave theory to estimate delays. The delay calculation can be combined with a user-specified priority list for vehicles or routes to generate waitingcosts per direction. This way, the decision-making goals of policy-makers can be translated into the priority weights of traffic modes. The resulting costs per directions are then used in the optimization by calculating the overall costs per direction for different signal plans. With the optimization, it is ensured that the overall waiting costs are minimized per intersection, by allocating green time to the direction with the highest waiting-costs.

It was found, that in these multimodal RTCS, little research was been done on the detection of spillback effects and providing priority when spillback effects occur. When spillback can be detected and controlled in a multimodal RTCS, unsafe situation are avoided or handled faster, so that less conflicts occur between different road users.

### **8.1.2. Design of spillback component**

For the thesis, it was chosen to design a spillback detection method that works with loop detectors in real-time applying shockwave theory. In most countries loop detectors are already placed in the road, furthermore, they are cheaper than using sensors that are able to measure areas around the intersection. However, in the test framework it was chosen to retrieve the queue length from the simulation software, due to a lack of time by the implementation complexity and calibration. Nevertheless, the queue length, measures cycle-by-cycle, and can be compared against a lane threshold in terms of meters. The approach detects either lane-blocking, spillover, or both. Another advantages is that the resulting shockwave profiles can also be used in the delay calculation for traffic light control.

With spillback detected, one of two designed control strategies were applied. The first designed strategy was a spillback penalty factor. The penalty factor multiplies the resulting costs with a weight to provide additional priority to the spillback direction. The second control strategy is an override strategy which interrupts the optimization algorithm to a phase plan in which the spillback direction receives green, neglecting the current phase plan.

### **8.2. Research question 2 – Effects on traffic and network performance**

The second research question was divided into two subquestions, namely the effect of spillback control strategies on the traffic and network performance of an isolated intersection. RQ2.2 investigated the effect of spillback control strategies on the traffic and network performance of a corridor.

### **8.2.1. Traffic and intersection performance isolated intersection**

In experiments, spillback was enforced on one side of the intersection. In general, the overall intersection delay increases when a spillback control strategy is applied in the traffic light control optimization. Regarding the height of the spillback penalty, only when the demand for disadvantaged directions is low enough, a higher spillback priority increases intersection performance. For higher demands, the height of the spillback penalty has no little to no difference in intersection performance.

In terms of delay, provision of additional priority for spillback also caused unbalanced allocation of green time. This occurred when multiple non-conflicting directions can receive green besides the spillback direction, and one of the directions clearly has higher demand. Therefore, individual vehicle wait exceptionally long at directions with low-demand, such as a left turn. When a conflicting busline is included in the traffic flow and the demand is low, the best overall results in delay are found with a small spillback penalty. With a higher demand, better are results not found with the application of a spillback penalty, however, it still reduced the delay of the spillback direction.

In terms of spillback duration, when the demand is higher a small spillback penalty (X5) leads to the best results. An increased penalty did decrease the duration of the spillback direction, however, it does not compensate for the increase in duration of other directions. When a conflicting busline is included, a low spillback penalty with a low demand has the best results for the spillback direction. The durations are more distributed among other directions when the spillback penalty is increased. With an increased demand and a low spillback penalty, improvements are still found for spillback directions in combination with a decrease in spillback for the bus direction. With increased penalty factors, traffic control leans into an advantage for spillback directions with higher delays for the bus and non-spillback directions.

The override strategy showed promising results on spillback duration; even with approaches competing in demand, benefits were shown for both the east and the west side. This is caused by the abrupt method, and neglecting safety constraints and phase structure. The benefits were paired very high delays for disadvantaged directions.

### **8.2.2. Traffic and network performance corridor**

In the corridor experiments, spillback was enforced on the north side of 3rd and 16th Street, halfway through the corridor. With a minor spillback penalty and a moderate vehicle flow, the overall network performance increases. Additionally, the network performance increased for all scenarios during the peak period. Except, when apart from spillback priority, also corridor priority was provided. Regarding the intersection performance, it can be said that the overall intersection delay decreases when the demand for conflicting directions is low, consequently, when the demand for conflicting directions is high, the gain in overall intersection performance is lost. Intersection performance is most often increased at the ends of the corridor. But also, the performance of the intersection downstream of intersection affected by spillback increased.

A spillback priority works well in combination with higher prioritized public transit, a penalty factor effectively decreases spillback durations and at the same time, minor increases in delay were found for the LRV. However, at one location in the corridor, passenger cars could get right-of-way with spillback priority over an LRV, for other directions the LRV travelled along the arterial. When also a bus was included in the traffic flow, lower delays and spillback durations were found for spillback directions. However, the effect was paired with higher delays for disadvantaged directions. Additionally, when buses are included in the spillback direction, spillback priority lead to practically absolute priority.

In terms of the height of the spillback penalty, a too-high priority negatively impacts the intersection and network performance. However, in some cases, a (high) spillback priority is necessary to provide enough priority so that a growing unsafe situation is avoided. Therefore, it is crucial to determine the intended use and the applicabilities of assigning spillback priority. The spillback priority promotes traffic and network performance when a short peak demand is expected. In more moderate to congested traffic situations, it is advised to look at whether the delay for disadvantaged directions is permitted. Furthermore, when an unsafe situation emerges, a very high spillback priority can be feasible, for example, when preemption is needed for an emergency vehicle and vehicles are spillt over at an intersection.

### **8.3. Discussion**

The researched literature and methods configuration were paired with several limitations, furthermore, the conclusions deduced from the performed experiments do not fully replicate real-life behaviour and are also paired with limitations. These are described in the next sections.

### **8.3.1. Designed methods**

The first limitation regards the utilized queuing model. The designed queuing model can accurately estimate queue length but the queue length is fully estimated after the cycle has been set and therefore is not predicted. Which influences the effectiveness of the spillback detection, as spillback is detected after it occurs. Consequently, the traffic control cleans up instead of avoiding the spillback effect. Additionally, the control method assigns additional priority when the queue length exceeds the lane capacity, but the delay road users experience is not into account in the allocation of green time. Regarding non-recurrent situations, the queue method does not take these situation into account, which can cause faulty traffic light control. For example, when a truck parks on the road for unloading cargo, the upstream traffic is blocked, and the allocated green time does not serve the intended use, but also the capacity of a lane decreases.

It is important to note that the 'best' spillback detection method (and control strategy) is dependent on the utilized traffic sensors. For example, when probe data or lane change behaviour is available in the otpimization of traffic light

control, more accurate control could be applied as more information is available.

### **8.3.2. Measured effects**

The results were obtained using a software-in-the-loop simulation. However, widely accepted for evaluating traffic simulations, it is paired with limitations due to the test configuration and the TFE algorithm. In this test configuration, the prediction component of TFE, a stochastic module in the system, makes identical predictions in every replication. The simulation outcomes varied more than ten percent in between with identical input, meaning ten runs were needed for a base case with a single departure profile for a meaningful average. Because of a long runtime, the random uncertainty, and the research goal of finding the effect of spillback control strategies, it was decided not to investigate multiple random seeds for the prediction variable in the TFE algorithm. Furthermore, the performed experiments assumed an exact and correct queue length retrieved from the simulation, consequently, the spillback detection method does not take into account sensor faults or misestimation of queue length. Regarding the single intersection network, an *earlywalk*-setting causes asynchronous green time. When north and southbound traffic both receive green at the isolated intersection, there is a lag of two seconds between the north and the south. This reduced the performance of the south side of the single intersections. However, this effect did not occur in the corridor.

Regarding the optimization module of TFE, the algorithm is set to evaluate every two seconds. The computational complexity increases when the demand is raised or the minimum green times are reduced for intersections. When the optimization exceeds the computational capacity to evaluate enough tasks few effects can occur. First, the simulation stalls until it is actively terminated, second, the simulation stalls for a while and then the run continues, causing massive delays and considerable differences between simulation runs. Third, a warning is provided for a short period of a few seconds to a minute, it is unclear if this has a large impact on the traffic control and the simulation outcomes.

Regarding the simulation environment and configuration, the improved performance at ends of the corridor can also be due to the vehicles leaving the network endlessly. Consequently, the network performance could be overestimated as no spillbacks occur from directions that leave the corridor. Furthermore, pedestrians are included in the simulation with fixed routes without adequate behaviour. In a few simulation runs, pedestrians got stuck on the side of crossings without a clear explanation. The effect is that in a few runs, the pedestrians do not cross the road and wait for a very long period before crossing a red light when no car is nearby. Therefore, no useful effect could be deduced from pedestrian behaviour. The behaviour of the drivers and the environment itself is limited in simulation, therefore, the effects may be underestimated or overestimated compared to a real-life setup. For example, minor disruptions such as a depart delay at intersection due to driver distraction, or more major disruptions such as lane closure upstream of an intersection due to roadworks or an incident.

Finally, the measured effects have been tested in undersaturated to moderate conditions, and demand profiles with lower demands at sidestreets and a higher demand along the mainstreets. Furthermore, the effects were tested with one set of vehicle priorities, providing priority to public transit. The measured effects should be seen in this light and could be different when an alternative demand profile is used, such as a higher demand for side streets or when vehicle priorities are configured differently. Consequently, additional spillback priority could have an increased influence on the delay performance of the LRV if public transit is assigned less priority in the network.

### **8.4. Recommendations**

Based on the findings and limitations of the study recommendations are done. First, recommendations are done for Technolution regarding policy implications and TFE model aspects, in which some are related to future research directions, also new research directions and topics are proposed.

### **8.4.1. Model and policy**

Recommendations are done for existing components of MM-Flow, but also for additional components. First, regarding the spillback detection method used in this thesis, in which spillback effects are coped with after they have occurred, it is advised to explore spillback prediction methods that avoid spillback occurrence instead of solving the occurred spillback. One of the approaches is calculating the probability of a spillover event, which has been developed by [Ramezani &](#page-84-0) [Geroliminis](#page-84-0) ([2015\)](#page-84-0). Second, in this thesis it was chosen to detect spillback by a queuing method. The queuing method used in the experiments is not applicable in real life, and the current queue model used in MM-Flow is insufficient for estimating queue lengths. If it is chosen in the future to include a queue method in the optimization, it can be useful to

look into a more sophisticated queue model so that queue lengths can be processed inside the TFE algorithm. Examples are the one suggested by [H. Liu et al.](#page-83-0) [\(2009](#page-83-0)), which has been extended with a Kalman filter by [Horvath M. & Tettamanti](#page-83-1) [T.](#page-83-1)([2021](#page-83-1)). Another advantage is that the generated shockwave profiles can also be used for delay calculation (see[Christofa](#page-82-0) [et al.](#page-82-0) ([2013](#page-82-0))). When sensors are used that can detect lane areas, real-time queue length measurement could be looked into, in which [Albiol et al.](#page-82-1) [\(2011](#page-82-1)) developed a pragmatic approach.

Regarding the spillback control strategy, the suggested height of the spillback priority depends on the situation and the intended use. A relative low weight not only promoted the traffic flow, but also increased the network performance on a corridor, however this effect is found with vehicle priorities that were configured so that the LRV received priority for the most part. Therefore, looking into the balance of assigning priority in the objective function is recommended. Furthermore, to look into assigning multiple weights of spillback priorities for different situations – for example, overflow on a short link needs a large spillback priority to control the spillover, and when a potential unsafe situation emerges benefits are also found when it is controlled as fast as possible. However, a high spillback priority at an ordinary intersection causes unnecessary long delays for disadvantaged directions. An alternative approach is to just promote the traffic flow by assigning additional priorities (movement-priority) for critical lanes instead of implementing a spillback penalty in the optimization. The current developed approach of the control strategy is to assign additional priority until the queue length is below a spillback detection queue threshold, therefore, it is advised to also look into controlling spillback effects based on the delays experienced by other cars. However, if the goal is to include the delays experienced by all types of traffic participants, the delays used in the optimization should be configured as personal delay instead of vehicle delay.

In terms of the optimization algorithm in the TFE, it is advised to look into optimization algorithms that search for a local optimum instead of evaluating all solutions so that TFE can handle more complexity, such as a more extensive network, a higher demand or additional computations. Furthermore, a test configuration without the ring and barrier structure could also be looked at so that the algorithm's performance can be tested with dynamic phasing, as the most pragmatic approach that neglects phase structure (override strategy) showed the highest reduction in spillback reduction.

Regarding the policy configuration of MM-Flow, it is recommended to gain insight into the effects of policy goals on the network and traffic performance. For example, the current set up showed the effects when the policy goal was to assign importance to public transport. But also other goals lead to other effects on the network performance, such as prioritizing vehicles based on their emissions

### **8.4.2. Future research**

During the thesis, other subjects and research directions emerged. In terms of spillback detection it is advised to further research the identification of non-recurrent congestion and the corresponding effects occurring from that congestion, such as the change in link capacity and handling that congestion in terms of traffic signal control. Second, regarding the priority provided to a spillback direction, it is advised to look further into how spillback should be prioritized in different types of traffic situations in a multimodal context. Additionally, more insight could be gained in the combination of different priorities, such as configuration of the vehicle priority in combination with additional spillback priority. What is the effect when only spillback priority is applied on the traffic and network performance. In terms of spillback control strategies, it is advised to look into the current strategy and vary with the weights given to the spillback penalty, dependent on specific traffic situations. Furthermore, the effect of the spillback control strategy with varying demand profiles. Furthermore, it is advised to further research other strategies to control spillback in a multimodal RTCS and the effect of that strategy on the traffic and network performance. Such approaches could include an alternative objective function when spillback is detected, or including downstream and upstream metering in the solve method. In terms of analyzing the effects of controlling spillback effects, it is advised to include speed contourplots to gain more insight per individual situation when a spillback control strategy is applied.

In terms of policy, it is recommended to gain insight into the effects of different combinations of priorities (vehicle, movement, and spillback) created to fulfill different policy goals, and the effects of these policy goals on the traffic and network performance. Altogether, when the policy goal is to reduce traffic fatalities to zero by 2024, it should come as no surprise that controlling an emerging unsafe situation as spillover at an upstream intersection should be highly prioritized in the traffic signal control.

# References

- <span id="page-82-1"></span>Albiol, A., Albiol, A., & Mossi, J. M. (2011). Video-based traffic queue length estimation. *Proceedings of the IEEE International Conference on Computer Vision*, 1928–1932. doi: 10.1109/ICCVW.2011.6130484
- <span id="page-82-6"></span>Amini, Z., Pedarsani, R., Skabardonis, A., & Varaiya, P. (2016, 12). Queue-length estimation using real-Time traffic data. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 1476–1481. doi: 10.1109/ITSC .2016.7795752
- <span id="page-82-4"></span>Auld, J., Hope, M., Ley, H., Sokolov, V., Xu, B., & Zhang, K. (2016, 3). POLARIS: Agent-based modeling framework development and implementation for integrated travel demand and network and operations simulations. *Transportation Research Part C: Emerging Technologies*, *64*, 101–116. doi: 10.1016/J.TRC.2015.07.017
- Baker, R. J., Collura, J., Dale, J. J., Head, L., Hemily, B., Ivanovic, M., … Smith, L. (2002, 9). An Overview of Transit Signal Priority.
- <span id="page-82-9"></span>Bhouri, N., Mayorano, F. J., Lotito, P. A., Salem, H. H., & Lebacque, J. P. (2015). Public Transport Priority for Multimodal Urban Traffic Control. *Bulgarian Academy Of Sciences Cybernetics And Information Technologies*, *15*(5). doi: 10.1515/cait-2015-0014
- Boillot, F., Blosseville, J. M., Lesort, J. B., Motyka, V., Papageorgiou, M., & Sellam, S. (1992). *Optimal Signal Control Of Urban Traffic Networks*. IEE Conf. Pub. 355.
- Boillot, F., Midenet, S., & Pierrelee, J. (2006). The real-time urban traffic control system CRONOS: Algorithm and experiments. doi: 10.1016/j.trc.2006.05.001
- Bretherton, R. (1990, 9). Scoot Urban Traffic Control System—Philosophy and Evaluation 1. *IFAC Proceedings Volumes*, *23*(2), 237–239. doi: 10.1016/S1474-6670(17)52676-2
- <span id="page-82-5"></span>Cao, X., Jiao, J., Zhang, Y., & Wang, X. (2019, 3). Left-Turn Spillback Probability Estimation in a Connected Vehicle Environment:. *https://doi.org/10.1177/0361198119837178*, *2673*(4), 753–761. Retrieved from [https://](https://journals.sagepub.com/doi/10.1177/0361198119837178) [journals.sagepub.com/doi/10.1177/0361198119837178](https://journals.sagepub.com/doi/10.1177/0361198119837178) doi: 10.1177/0361198119837178
- <span id="page-82-8"></span>Chen, Y. Y., & Chang, G. L. (2014). A macroscopic signal optimization model for arterials under heavy mixed traffic flows. *IEEE Transactions on Intelligent Transportation Systems*, *15*(2), 805–817. doi: 10.1109/TITS.2013.2289961
- <span id="page-82-3"></span>Christofa, E. (2012). Traffic Signal Optimization with Transit Priority: A Person-based Approach.
- <span id="page-82-2"></span>Christofa, E., Ampountolas, K., & Skabardonis, A. (2016, 5). Arterial traffic signal optimization: A person-based approach. *Transportation Research Part C: Emerging Technologies*, *66*, 27–47. doi: 10.1016/J.TRC.2015.11.009
- <span id="page-82-0"></span>Christofa, E., Argote, J., & Skabardonis, A. (2013, 1). Arterial Queue Spillback Detection and Signal Control Based on Connected Vehicle Technology:. *https://doi.org/10.3141/2356-08*(2356), 61–70. Retrieved from [https://](https://journals.sagepub.com/doi/10.3141/2356-08) [journals.sagepub.com/doi/10.3141/2356-08](https://journals.sagepub.com/doi/10.3141/2356-08) doi: 10.3141/2356-08
- Daganzo, C. F. (1995, 4). The cell transmission model, part II: Network traffic. *Transportation Research Part B: Methodological*, *29*(2), 79–93. doi: 10.1016/0191-2615(94)00022-R
- <span id="page-82-7"></span>Di Gangi, M., Cantarella, G. E., Pace, R. D., & Memoli, S. (2016, 5). Network traffic control based on a mesoscopic dynamic flow model. *Transportation Research Part C: Emerging Technologies*, *66*, 3–26. doi: 10.1016/J.TRC.2015 .10.002
- Ding, J., He, Q., Head, L., Saleem, F., & Wu, W. (2013, 1). *Development and Testing of Priority Control System in Connected Vehicle Environment* (Tech. Rep.). Washington: Transportation Research Board. Retrieved from <https://trid.trb.org/view/1241792>
- Fietserbond. (2022, Apr). *Verkeerslichten.* Retrieved from [https://www.fietsersbond.nl/ons-werk/](https://www.fietsersbond.nl/ons-werk/infrastructuur/verkeerslichten/) [infrastructuur/verkeerslichten/](https://www.fietsersbond.nl/ons-werk/infrastructuur/verkeerslichten/)
- <span id="page-83-5"></span>Flötteröd, G., & Osorio, C. (2017, 8). Stochastic network link transmission model. *Transportation Research Part B: Methodological*, *102*, 180–209. Retrieved from <http://dx.doi.org/10.1016/j.trb.2017.04.009> doi: 10.1016/J.TRB.2017.04.009
- Gartner, N. H., Pooran, F. J., & Andrews, C. M. (2001). Implementation of the OPAC adaptive control strategy in a traffic signal network. In *Ieee conference on intelligent transportation systems, proceedings, itsc* (pp. 195–200). doi: 10.1109/itsc.2001.948655
- <span id="page-83-6"></span>Han, K., & Gayah, V. V. (2015, 7). Continuum signalized junction model for dynamic traffic networks: Offset, spillback, and multiple signal phases. *Transportation Research Part B: Methodological*, *77*, 213–239. doi: 10.1016/J.TRB.2015 .03.005
- <span id="page-83-2"></span>He, Q., Head, K. L., & Ding, J. (2014). Multi-modal traffic signal control with priority, signal actuation and coordination. *Transportation Research Part C: Emerging Technologies*, *46*, 65–82. doi: 10.1016/j.trc.2014.05.001
- Hoogendoorn, S., & Knoop, V. (2016a). *Fundamental diagrams*. Retrieved from [https://ocw.tudelft.nl/](https://ocw.tudelft.nl/wp-content/uploads/Chapter-4.-Fundamental-diagrams.pdf) [wp-content/uploads/Chapter-4.-Fundamental-diagrams.pdf](https://ocw.tudelft.nl/wp-content/uploads/Chapter-4.-Fundamental-diagrams.pdf)
- Hoogendoorn, S., & Knoop, V. (2016b). *Shock wave analysis*. Retrieved from [https://ocw.tudelft.nl/](https://ocw.tudelft.nl/wp-content/uploads/Chapter-8.-Shock-wave-analysis.pdf) [wp-content/uploads/Chapter-8.-Shock-wave-analysis.pdf](https://ocw.tudelft.nl/wp-content/uploads/Chapter-8.-Shock-wave-analysis.pdf)
- <span id="page-83-1"></span>Horvath M., & Tettamanti T. (2021). *Real-time Queue Length Estimation Applying Shockwave Theory at Urban Signalized Intersections.* Retrieved from <https://pp.bme.hu/ci/article/view/17022/9134> doi: 10.3311/PPci.17022
- Hounsell, N., & Shrestha, B. (2005, 1). AVL based Bus Priority at Traffic Signals. *European Journal of Transport and Infrastructure Research*, *5*(1), 13–29. Retrieved from [https://journals.open.tudelft.nl/ejtir/](https://journals.open.tudelft.nl/ejtir/article/view/4330) [article/view/4330](https://journals.open.tudelft.nl/ejtir/article/view/4330) doi: 10.18757/EJTIR.2005.5.1.4330
- <span id="page-83-3"></span>Jovanović, A., Kukić, K., & Stevanović, A. (2021, 4). A fuzzy logic simulation model for controlling an oversaturated diverge diamond interchange and ramp metering system. *Mathematics and Computers in Simulation*, *182*, 165–181. doi: 10.1016/J.MATCOM.2020.10.023
- Light, R. A. (2017). Mosquitto: server and client implementation of the mqtt protocol. *Journal of Open Source Software*, *2*(13), 265. Retrieved from <https://doi.org/10.21105/joss.00265> doi: 10.21105/joss.00265
- Lighthill, M. J., & Whitham, G. B. (1955, 5). On kinematic waves II. A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, *229*(1178), 317–345. Retrieved from<https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1955.0089> doi: 10.1098/RSPA.1955.0089
- <span id="page-83-0"></span>Liu, H., Wu, X., Ma, W., & Hu, H. (2009, 8). Real-time queue length estimation for congested signalized intersections. *Transportation Research Part C: Emerging Technologies*, *17*(4), 412–427. doi: 10.1016/J.TRC.2009.02.003
- <span id="page-83-4"></span>Liu, S., Li, Y., Qiu, Y., Zhang, B., Qiu, S., & Liu, X. (2020, 10). Signal Timing Optimization Algorithm for an Intersection Connected with an Urban Expressway. *Arabian Journal for Science and Engineering*, *45*(10), 8667–8684. Retrieved from <https://link.springer.com/article/10.1007/s13369-020-04835-6> doi: 10.1007/S13369-020-04835-6/FIGURES/10
- <span id="page-83-7"></span>Liu, Y., & Chang, G. L. (2011, 2). An arterial signal optimization model for intersections experiencing queue spillback and lane blockage. *Transportation Research Part C: Emerging Technologies*, *19*(1), 130–144. doi: 10.1016/J.TRC .2010.04.005
- Liu, Y., Guo, J., & Wang, Y. (2018). Vertical and Horizontal Queue Models for Oversaturated Signal Intersections with Quasi-Real-Time Reconstruction of Deterministic and Shockwave Queueing Profiles Using Limited Mobile Sensing Data. Retrieved from <https://doi.org/10.1155/2018/6986198> doi: 10.1155/2018/6986198
- Lopez, P. A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., … Wießner, E. (2018). Microscopic traffic simulation using sumo. In *The 21st ieee international conference on intelligent transportation systems.* IEEE. Retrieved from <https://elib.dlr.de/124092/>
- Ma, D., Luo, X., Jin, S., Wang, D., & Guo, W. (2018, 9). Estimating Maximum Queue Length for Traffic Lane Groups Using Travel Times from Video-Imaging Data. *IEEE Intelligent Transportation Systems Magazine*, *10*(3), 123–134. doi: 10.1109/MITS.2018.2842047
- Mauro, V., & Di Taranto, C. (1990, 9). UTOPIA. *IFAC Proceedings Volumes*, *23*(2), 245–252. Retrieved from <https://linkinghub.elsevier.com/retrieve/pii/S1474667017526786> doi: 10.1016/ S1474-6670(17)52678-6
- <span id="page-84-1"></span>Mein, E., Van Koningsbruggen, P., & Van Hoekstra, P. (2022). *Multi-modal Micro-management Agent-based Corridor Optimization* (Tech. Rep.).
- Mirchandani, P., & Head, L. (2001). A real-time traffic signal control system: Architecture, algorithms, and analysis. *Transportation Research Part C: Emerging Technologies*, *9*(6), 415–432. Retrieved from [www.elsevier.com/](www.elsevier.com/locate/trc) [locate/trc](www.elsevier.com/locate/trc) doi: 10.1016/S0968-090X(00)00047-4
- <span id="page-84-3"></span>Mohajerpoor, R., & Cai, C. (2020, 7). A Traffic Signal Control Strategy to Avoid Spillback on Short Links. *Proceedings of the American Control Conference*, *2020-July*, 1167–1172. doi: 10.23919/ACC45564.2020.9147494
- <span id="page-84-2"></span>Noaeen, M., Mohajerpoor, R., H. Far, B., & Ramezani, M. (2021, 12). Real-time decentralized traffic signal control for congested urban networks considering queue spillbacks. *Transportation Research Part C: Emerging Technologies*, *133*, 103407. doi: 10.1016/J.TRC.2021.103407
- Office of Operations. (2021). *Traffic signal timing manual.* US Department of Transportation. Retrieved from <https://ops.fhwa.dot.gov/publications/fhwahop08024/chapter4.htm>
- Perez-Montesinos, J., Dixon, M. P., & Kyte, M. (2011, 1). Detection of Stop Bar Traffic Flow State:. *https://doi.org/10.3141/2259-12*, *2259*(2259), 132–140. Retrieved from [https://journals.sagepub](https://journals.sagepub.com/doi/abs/10.3141/2259-12?casa_token=Yj6zzZSrpCAAAAAA%3AMXhvdJgO3NxVTD-zc4yw4JaB3RvdstLHHv-iGA3mQ6HRRWF5gNaHmGVgQipwntCY1uGSW_3BNAqy) [.com/doi/abs/10.3141/2259-12?casa\\_token=Yj6zzZSrpCAAAAAA%3AMXhvdJgO3NxVTD](https://journals.sagepub.com/doi/abs/10.3141/2259-12?casa_token=Yj6zzZSrpCAAAAAA%3AMXhvdJgO3NxVTD-zc4yw4JaB3RvdstLHHv-iGA3mQ6HRRWF5gNaHmGVgQipwntCY1uGSW_3BNAqy) [-zc4yw4JaB3RvdstLHHv-iGA3mQ6HRRWF5gNaHmGVgQipwntCY1uGSW\\_3BNAqy](https://journals.sagepub.com/doi/abs/10.3141/2259-12?casa_token=Yj6zzZSrpCAAAAAA%3AMXhvdJgO3NxVTD-zc4yw4JaB3RvdstLHHv-iGA3mQ6HRRWF5gNaHmGVgQipwntCY1uGSW_3BNAqy) doi: 10.3141/2259-12
- <span id="page-84-4"></span>Ramezani, M., De Lamberterie, N., Skabardonis, A., & Geroliminis, N. (2016). A link partitioning approach for realtime control of queue spillbacks on congested arterials. Retrieved from [http://www.tandfonline.com/](http://www.tandfonline.com/action/journalInformation?journalCode=ttrb20) [action/journalInformation?journalCode=ttrb20](http://www.tandfonline.com/action/journalInformation?journalCode=ttrb20) doi: 10.1080/21680566.2016.1142399
- Ramezani, M., & Geroliminis, N. (2013). Exploiting probe data to estimate the queue profile in urban networks. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 1817–1822. doi: 10.1109/ITSC.2013.6728492
- <span id="page-84-0"></span>Ramezani, M., & Geroliminis, N. (2015, 6). Queue Profile Estimation in Congested Urban Networks with Probe Data. *Computer-Aided Civil and Infrastructure Engineering*, *30*(6), 414–432. Retrieved from [https://onlinelibrary.wiley.com/doi/full/10.1111/mice.12095https://](https://onlinelibrary.wiley.com/doi/full/10.1111/mice.12095https://onlinelibrary.wiley.com/doi/abs/10.1111/mice.12095https://onlinelibrary.wiley.com/doi/10.1111/mice.12095) [onlinelibrary.wiley.com/doi/abs/10.1111/mice.12095https://onlinelibrary](https://onlinelibrary.wiley.com/doi/full/10.1111/mice.12095https://onlinelibrary.wiley.com/doi/abs/10.1111/mice.12095https://onlinelibrary.wiley.com/doi/10.1111/mice.12095) [.wiley.com/doi/10.1111/mice.12095](https://onlinelibrary.wiley.com/doi/full/10.1111/mice.12095https://onlinelibrary.wiley.com/doi/abs/10.1111/mice.12095https://onlinelibrary.wiley.com/doi/10.1111/mice.12095) doi: 10.1111/MICE.12095
- <span id="page-84-6"></span>Ren, Y., Wang, Y., Yu, G., Liu, H., & Xiao, L. (2017, 6). An Adaptive Signal Control Scheme to Prevent Intersection Traffic Blockage. *IEEE Transactions on Intelligent Transportation Systems*, *18*(6), 1519–1528. doi: 10.1109/TITS .2016.2609917
- Robertson, D. I., & Bretherton, R. D. (1991). Optimizing Networks of Traffic Signals in Real Time-The SCOOT Method. *IEEE Transactions On Vehicular Technology*, *40*(1), 11–15. Retrieved from [https://ieeexplore](https://ieeexplore.ieee.org/document/69966) [.ieee.org/document/69966](https://ieeexplore.ieee.org/document/69966) doi: 10.1109/25.69966
- San Francisco County Transportation Authority. (2018). *Vision Zero.* Retrieved from [https://www.sfcta](https://www.sfcta.org/policies/vision-zero) [.org/policies/vision-zero](https://www.sfcta.org/policies/vision-zero)
- <span id="page-84-5"></span>Sharma, A., Bullock, D. M., Bonneson, J. A., Sharma, A. ., & Bullock, D. M. . (2007). Input-Output and Hybrid Techniques for Real-Time Prediction of Input-Output and Hybrid Techniques for Real-Time Prediction of Delay and Maximum Queue Length at Signalized Intersections. Retrieved from [https://digitalcommons.unl](https://digitalcommons.unl.edu/civilengfacpub.17.https://digitalcommons.unl.edu/civilengfacpub/17) [.edu/civilengfacpub.17.https://digitalcommons.unl.edu/civilengfacpub/17](https://digitalcommons.unl.edu/civilengfacpub.17.https://digitalcommons.unl.edu/civilengfacpub/17) doi: 10.3141/2035-08
- Sims, A. G., & Dobinson, K. W. (1980). The Sydney Coordinated Adaptive Traffic (SCAT) System Philosophy and Benefits. *IEEE Transactions on Vehicular Technology*, *29*(2), 130–137. doi: 10.1109/T-VT.1980.23833
- Stevanovic, A., Abdel-Rahim, A., Zlatkovic, M., & Amin, E. (2009). Microscopic modeling of traffic signal operations: Comparative evaluation of hardware-in-the-loop and software-in-the-loop simulations. *Transportation Research Record*(2128), 143–151. doi: 10.3141/2128-15
- Stevanovic, A., Kergaye, C., & Martin, P. T. (2009, 1). SCOOT and SCATS: A Closer Look into Their Operations. In *Transportation research board annual meeting.* Retrieved from [https://www.researchgate.net/](https://www.researchgate.net/publication/274137098) [publication/274137098](https://www.researchgate.net/publication/274137098)
- <span id="page-85-2"></span>Stevanovic, A., Zlatkovic, M., Stevanovic, J., Kergaye, C., Ostojic, M., & Tasic, I. (2015). Multimodal Traffic Control for Large Urban Networks with Special Priority for Light Rail Transit. In *Proceedings of the 94th annual meeting of the transportation.* Retrieved from <https://www.researchgate.net/publication/274137297>
- <span id="page-85-3"></span>Stevanovic, J. (2011). Traffic control optimization for multi-modal operations in a large-scale urban network.
- Stewart, T. (2022). *Overview of motor vehicle crashes in 2020 (report no. dot hs813266)*. Retrieved from [https://](https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813266) [crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813266](https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813266)
- SUMO. (2022). *Using additional polygons and pois within the simulation*. Eclipse. Retrieved from [https://](https://sumo.dlr.de/docs/Simulation/Shapes.html) [sumo.dlr.de/docs/Simulation/Shapes.html](https://sumo.dlr.de/docs/Simulation/Shapes.html)
- <span id="page-85-4"></span>Tilg, G., Ambühl, L., Batista, S., Menendez, M., & Busch, F. (2021, 8). On the application of variational theory to urban networks. *Transportation Research Part B: Methodological*, *150*, 435–456. doi: 10.1016/J.TRB.2021.06.019
- Umair, M., Farooq, M. U., Raza, R. H., Chen, Q., & Abdulhai, B. (2021, 10). Efficient video-based vehicle queue length estimation using computer vision and deep learning for an urban traffic scenario. *Processes*, *9*(10). doi: 10.3390/ pr9101786
- <span id="page-85-0"></span>Van Katwijk, R. (2008). *Multi-Agent Look-Ahead Traffic-Adaptive Control* (Tech. Rep.). Retrieved from [http://](http://resolver.tudelft.nl/uuid:92a019ab-d5c7-4b33-8643-1548014a3a3b) [resolver.tudelft.nl/uuid:92a019ab-d5c7-4b33-8643-1548014a3a3b](http://resolver.tudelft.nl/uuid:92a019ab-d5c7-4b33-8643-1548014a3a3b)
- Wahlstedt, J. (2013). Evaluation of the two self-optimising traffic signal systems Utopia/Spot and ImFlow, and comparison with existing signal control in Stockholm, Sweden. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 1541–1546. doi: 10.1109/ITSC.2013.6728449
- Wang, Z., Cai, Q., Wu, B., Zheng, L., & Wang, Y. (2016). Shockwave-based queue estimation approach for undersaturated and oversaturated signalized intersections using multi-source detection data. Retrieved from [https://www](https://www.tandfonline.com/action/journalInformation?journalCode=gits20) [.tandfonline.com/action/journalInformation?journalCode=gits20](https://www.tandfonline.com/action/journalInformation?journalCode=gits20) doi: 10.1080/ 15472450.2016.1254046
- <span id="page-85-5"></span>Wong, C. k., & Lee, Y. y. (2020, 8). Lane-Based Traffic Signal Simulation and Optimization for Preventing Overflow. *Mathematics 2020, Vol. 8, Page 1368*, *8*(8), 1368. Retrieved from [https://www.mdpi.com/2227-7390/](https://www.mdpi.com/2227-7390/8/8/1368/htmhttps://www.mdpi.com/2227-7390/8/8/1368) [8/8/1368/htmhttps://www.mdpi.com/2227-7390/8/8/1368](https://www.mdpi.com/2227-7390/8/8/1368/htmhttps://www.mdpi.com/2227-7390/8/8/1368) doi: 10.3390/MATH8081368
- Wu, J., Xu, H., Zhang, Y., Tian, Y., & Song, X. (2020, 4). Real-Time Queue Length Detection with Roadside LiDAR Data. *Sensors 2020, Vol. 20, Page 2342*, *20*(8), 2342. Retrieved from[https://www.mdpi.com/1424-8220/](https://www.mdpi.com/1424-8220/20/8/2342/htmhttps://www.mdpi.com/1424-8220/20/8/2342) [20/8/2342/htmhttps://www.mdpi.com/1424-8220/20/8/2342](https://www.mdpi.com/1424-8220/20/8/2342/htmhttps://www.mdpi.com/1424-8220/20/8/2342) doi: 10.3390/S20082342
- <span id="page-85-6"></span>Wu, X., Liu, H. X., & Gettman, D. (2010, 8). Identification of oversaturated intersections using high-resolution traffic signal data. *Transportation Research Part C: Emerging Technologies*, *18*(4), 626–638. doi: 10.1016/J.TRC.2010.01 .003
- Xie, X.-F., Smith, S. F., & Barlow, G. J. (2015, October 13). *Smart and scalable urban signal networks: methods and systems for adaptive traffic signal control.* Google Patents. (US Patent 9,159,229)
- <span id="page-85-7"></span>Xie, X.-F., Smith, S. F., Barlow, G. J., & Chen, T.-W. (2014). Coping with Real-World Challenges in Real-Time Urban Traffic Control. In *Compendium of papers of the 93rd annual meeting of the transportation research board* (pp. 1–15). Retrieved from <http://www.wiomax.com/traffic>
- <span id="page-85-1"></span>Yagar, S., & Han, B. (1994, 8). A procedure for real-time signal control that considers transit interference and priority. *Transportation Research Part B: Methodological*, *28*(4), 315–331. doi: 10.1016/0191-2615(94)90004-3
- <span id="page-86-3"></span>Yang, X., Cheng, Y., & Chang, G. L. (2014). Real-time traffic queue length estimation at the freeway off-ramp using dual-zone detectors. *21stWorld Congress on Intelligent Transport Systems, ITSWC 2014: Reinventing Transportation in Our Connected World*.
- Zargari, S. A., Dehghani, N., Mirzahossein, H., & Hamedi, M. (2016, 3). Improving scats operation during congestion periods using internal/external traffic metering strategy. *Promet - Traffic - Traffico*, *28*(1), 41–47. doi: 10.7307/ PTT.V28I1.1718
- <span id="page-86-0"></span>Zeng, X., Sun, X., Zhang, Y., & Quadrifoglio, L. (2019, 4). Person-Based Adaptive Priority Signal Control with Connected-Vehicle Information:. *https://doi.org/10.3141/2487-07*, *2487*, 78–87. Retrieved from [https://journals.sagepub.com/doi/abs/10.3141/2487-07?casa](https://journals.sagepub.com/doi/abs/10.3141/2487-07?casa_token=bxoRbdLtdsoAAAAA%3A5mCzHlOYBBFj6_w_0340QXVLa4Z7jdb8iV7zOXy_MXdK0QrRySnpv3CGrpB004JkwnOMGNrg-JVu) [\\_token=bxoRbdLtdsoAAAAA%3A5mCzHlOYBBFj6\\_w\\_0340QXVLa4Z7jdb8iV7zOXy](https://journals.sagepub.com/doi/abs/10.3141/2487-07?casa_token=bxoRbdLtdsoAAAAA%3A5mCzHlOYBBFj6_w_0340QXVLa4Z7jdb8iV7zOXy_MXdK0QrRySnpv3CGrpB004JkwnOMGNrg-JVu) [\\_MXdK0QrRySnpv3CGrpB004JkwnOMGNrg-JVu](https://journals.sagepub.com/doi/abs/10.3141/2487-07?casa_token=bxoRbdLtdsoAAAAA%3A5mCzHlOYBBFj6_w_0340QXVLa4Z7jdb8iV7zOXy_MXdK0QrRySnpv3CGrpB004JkwnOMGNrg-JVu) doi: 10.3141/2487-07
- <span id="page-86-1"></span>Zhang, H., Liu, H. X., Chen, P., Yu, G., & Wang, Y. (2020, 8). Cycle-Based End of Queue Estimation at Signalized Intersections Using Low-Penetration-Rate Vehicle Trajectories. *IEEE Transactions on Intelligent Transportation Systems*, *21*(8), 3257–3272. doi: 10.1109/TITS.2019.2925111
- <span id="page-86-4"></span>Zhang, Y., & Tong, J. (2008, 1). Modeling Left-Turn Blockage and Capacity at Signalized Intersection with Short Left-Turn Bay. *https://doi.org/10.3141/2071-09*(2071), 71–76. Retrieved from [https://journals.sagepub](https://journals.sagepub.com/doi/10.3141/2071-09) [.com/doi/10.3141/2071-09](https://journals.sagepub.com/doi/10.3141/2071-09) doi: 10.3141/2071-09
- <span id="page-86-2"></span>Zheng, F., Van Zuylen, H., & Liu, X. (2017, 8). A methodological framework of travel time distribution estimation for urban signalized arterial roads. *Transportation Science*, *51*(3), 893–917. doi: 10.1287/TRSC.2016.0718

# $\overline{\mathcal{A}}$

# Literature review tables

Table A.1: Overview of considered papers





Table A.2: Overview of researched paper literature Table A.2: Overview of researched paper literature **SB type**: LB (lane-blocking), SO (Spillover). Detection: LC (Link capacity), OSI (Oversaturation Severity Index), OT (Queue Threshold), Prob() (Probability of spillback type), SD (Speed detection).<br>**SB contro**l: OF (Objec **SB type**: LB (lane-blocking), SO (Spillover). Detection: LC (Link capacity), OSI (Oversaturation Severity Index), OT (Queue Threshold), Prob() (Probability of spillback type), SD (Speed detection).<br>ED contract Chistoriati **SB control**: OF (Objective function), PinOF (Penalty in objective function), M- (Metering from US and/or DS direction), US (Upstream), DS (Downstream), SS (Sidestreet)



Figure A.1: Residual queue shockwave profile

### **A.1. Queue length estimation**

When the traffic light control allocates insufficient green time, or when the traffic conditions are oversaturated, a residual queue  $L^n_{min}$  could form at time instant  $T^n_{min}$ . Using shockwave speed  $\omega_4$  and the queue discharging process between  $T_{max}^n$  and  $T_{min}^n$ , defined by  $\omega_2$ , this can be estimated. The maximum queue length and the time instant it occurs is already estimated (see chapter 4).

$$
L_{min}^n = \left(\frac{L_{max}^n}{\omega_2} + T_{max}^n - T_r^{n+1}\right) / \left(\frac{1}{\omega_2} + \frac{1}{\omega_4}\right)
$$
 (A.1)

$$
T_{\min}^n = T_t^{n+1} + L_{\min}^n / \omega_4
$$
 (A.2)

The last queue state is derivable from the maximum queue length, the time it occurs, and the time instant  $T_A$ .  $\omega_3$  can be estimated, assuming a constant shockwave speed, by:

$$
\omega_3 = (L_{max}^n - L_d)/(T_{max}^n - T_A) \tag{A.3}
$$

#### **A.1.1. Estimation of shockwave variables**

For the estimation of  $\omega_1$ , the current inflow  $q_i$  and current density  $k_i$  should be known, which is possible by traffic state estimation of the upstream detector, as the vehicles pass the sensor with the state  $q_i$  and  $k_i.$  For  $\omega_3,$  only the jam density should be configured; after that, it is possible to estimate  $\omega_3$ , with **??**. The advance detectors deliver data to estimate  $\omega_2$  and  $\omega_4$ . This way,  $\omega_4$  is calculated using the distance between the stopbar and the advance sensor:  $L_d$ , and the time difference between the green time start  $T_g^n$  and discharge wave reaching advance sensor:  $T_s$ . Another option would be with an assumed saturation flow  $q_max$ , jam density  $k_j$ , and critical density  $k_c$ .

$$
\omega_4 = L_d / (T_s - T_g^n) \tag{A.4}
$$

Shock wave speed  $\omega_2$  is estimated by identifying the states maximum flow  $q_{max}$ , critical density  $k_m$ , and the current inflow  $q_i$  and current density  $k_i.$  However, to estimate  $\omega_2,$  event-based data should be available to record the detector occupancy time. The detector occupancy time  $t_{o,i}$ , and the sum of the vehicle length plus the detector length  $l_e$  are used to estimate individual vehicle speed *u<sup>i</sup>* ([Equation A.5\)](#page-90-0). Then, with the number of vehicles *n* known, the sum of the individual vehicle speed *u<sup>i</sup>* , divided by the number of vehicles, calculates the space mean speed *uspace*, according to [Equation A.6.](#page-90-1)

These conditions can be estimated using the equations [Equation A.5](#page-90-0) to [Equation A.8](#page-90-2) to calculate shockwave  $\omega_2$ .

Now, the sum of headways *h<sup>i</sup>* of the vehicles estimates the average flow. The sum of the detector time occupancy *o* of vehicle *i*: *to,i* and the time gap *g* of vehicle *i* with its predecessor: *to,i*, divided by the number of cars [\(Equation A.7\)](#page-90-3) calculates the headway per vehicle. Lastly, the average density is calculated using the relationship between the density, flow, and speed  $(q = ku)$ , as shown in [Equation A.8](#page-90-2).

<span id="page-90-0"></span>
$$
u_i = l_e / t_{o,i} \tag{A.5}
$$

<span id="page-90-1"></span>
$$
u_{space} = 1/(\frac{1}{n} \sum_{i=1}^{n} \frac{1}{u_i})
$$
 (A.6)

<span id="page-90-3"></span>
$$
q = 1/(\frac{1}{n} \sum_{i=1}^{n} (h_i))
$$
 (A.7)

<span id="page-90-2"></span>
$$
k = q/u_{space}
$$
 (A.8)

### **A.1.2. Estimation points A, B, and C**

Point A, indicated by  $T_A$ , is the time instant that queue shockwave  $\omega_3$  propagates backwards to the location of the advance detector. Between the beginning of red in cycle *n* and *TA*, the vehicles travel over the loop detector with the traffic state  $(q_i,k_i)$ . After passing the loop detector, between  $T_A$  and  $T_{max}^n$ , no flow is possible due to the jam density:  $(0, k_i)$ . Point A identifies whether there is a long queue; when point A does not exist, the queue does not propagate further than the advance detector or outside the lane area detector range. A detector occupancy threshold is used to detect point A, or in the case of a lane area detector, the sensor range that still provides an accurate queue length.

B is the time instant  $T_b$  in which the queue discharge shockwave passed the advance detector with a shockwave speed of  $\omega_4$ . The traffic state is also a jammed state of  $(0,k_j)$  between the green start  $T_g^n$  and the time instant  $T_b$ . The vehicles discharge according to the maximum flow assumption of  $(q_{max}, k_c)$ . Because the traffic density is 0 after  $T_g^n$  and in advance of  $T_b,$  the time occupancy of the detector is high. After that, the queue discharge over the detector or inside the lane area range, and the occupancy time and the time gap between vehicles drop significantly. It is advisable to use sensor data to estimate the (*qmax, kc*), as the saturation flow could decrease due to busy downstream intersections.

 $T_C$  is the time instant in which the tail of the queue travels over the detector. Point C estimates the maximum queue length, combined with the queue discharge shock wave  $\omega_4$  to describe the queuing process.  $\omega_2$  is the shockwave between the arrival flow state  $(q_i,k_i)$  and  $(q_{max},k_c)$  and describes the vehicle discharge in which occurs between  $T_{max}^n$  and  $T_C.$  After the wave propagates to point C, the traffic state changes to  $(q_i,k_i)$ , which means that the discharge rate at the advance detector location is less than the maximum flow. The time gap between two successive vehicles is sensitive to the change in traffic state: before  $T_C$ , the time gap is smaller than 2.5 seconds with a small variance. After that, the gaps increase in combination with the variance. In addition, a time lag usually occurs between the queue discharge flow and newly arrived traffic([H. Liu et al.,](#page-83-0) [2009](#page-83-0)). Based on their observations, when the time gap exceeds 2.5 seconds, the queue's end propagates forward, the detector line is reached, and  $T_C$  is identified. For verification, the system could check for multiple occasions where the time gap exceeds 2.5 seconds.

# B

# Network configuration measurements

# **B.1. Isolated intersection**

**B.1.1. Network dimensions**



Figure B.1: Network model

Table B.1: Table with network specifications



# **B.1.2. Traffic light control specification**

Table B.2: Phase durations isolated intersection



# **B.2. Corridor**

## **B.2.1. Network dimensions**



Table B.3: Overview of edge measurements between intersection nodes



# **B.2.2. Traffic light control specification**

Table B.4: Traffic light time configuration (part 1)



Table B.5: Traffic light time configuration (part 2)



# **B.2.3. Turning fractions**

Table B.6: Turning fractions corridor



East.3rd\_18th 3rd\_18th.3rd\_19th.1 33.33333333

East.3rd\_18th 3rd\_18th.West 33.33333333 East.3rd\_19th 3rd\_19th.3rd\_18th.1 33.33333333 East.3rd\_19th 3rd\_19th.3rd\_20th.1 33.33333333 East.3rd\_19th 3rd\_19th.West 33.33333333 East.3rd\_20th 3rd\_20th.3rd\_19th.1 33.33333333 East.3rd\_20th 3rd\_20th.South.1 33.33333333 East.3rd\_20th 3rd\_20th.West 33.33333333





# **B.2.4. Demand profile**

#### Table B.7: Resulting demand profile



# $\bigcirc$

# TFE results isolated intersection cars

# **C.1. TFE Demand factor 1**



Figure C.1: Average total delay in [s]

Table C.1: Average total delay per scenario [s]





### **C.1.1. Resulting graphs of KPIs**

Figure C.2: Spillback duration in [s]

### **C.1.2. Base case scenario**

Table C.2: Overview KPIs of basecase





![](_page_102_Figure_2.jpeg)

# **C.1.3. Result X5**

Table C.3: Overview KPIs

![](_page_103_Picture_199.jpeg)

## **C.1.4. Result X10**

Table C.4: Overview KPIs

![](_page_103_Picture_200.jpeg)

### **C.1.5. Result X1000**

Table C.5: Overview KPIs

![](_page_104_Picture_199.jpeg)

### **C.1.6. Result override**

Table C.6: Overview KPIs

![](_page_104_Picture_200.jpeg)

![](_page_105_Figure_1.jpeg)

# **C.2. TFE Demand factor 1.5**

Figure C.5: Average total delay in [s]

Table C.7: Average total delay per scenario [s]

![](_page_105_Picture_102.jpeg)

### **C.2.1. Resulting graphs of KPIs**

![](_page_105_Figure_7.jpeg)

Figure C.6: Spillback duration in [s]

![](_page_106_Figure_1.jpeg)

### **C.2.2. Basecase**

Table C.8: Overview KPIs of basecase

![](_page_107_Picture_196.jpeg)

## **C.2.3. Result X5**

Table C.9: Overview KPIs

![](_page_107_Picture_197.jpeg)
#### **C.2.4. Result X10**

Table C.10: Overview KPIs



#### **C.2.5. Result X1000**

Table C.11: Overview KPIs



#### **C.2.6. Result override**

Table C.12: Overview KPIs



#### **C.3. TFE alternative turning fractions east**



Figure C.9: Average total delay in [s]

<b>Direction</b>	Mean	Std.dev	t-value	dF	Significant?
<b>Baseccase</b>	240.280	5.899	0.000	5.793	Nο
Result x5	241.724	7.111	$-1.563$	6.939	(p < 0.25)
Result x10	243.732	8.989	$-3.210$	7.996	Yes: $(p < 0.05)$
Result x1000	242.987	8.469	-2.622	7.778	Yes: $(p < 0.05)$
Result override	252.992	10.497	$-10.557$	8.418	Yes: $(p < 0.05)$

Table C.13: Average total delay per scenario [s]

#### **C.3.1. Resulting graphs of KPIs**



Figure C.10: Duration of spillback in [s]

#### **C.3.2. Base case**

Table C.14: Overview KPIs of basecase





#### **C.3.3. Result X5**

Table C.15: Overview KPIs



#### **C.3.4. Result X10**

Table C.16: Overview KPIs



#### **C.3.5. Result X1000**

Table C.17: Overview KPIs



#### **C.3.6. Result override**

Table C.18: Overview KPIs



# D

# TFE results isolated intersection with bus

#### **D.1. TFE with bus, experiment demand factor 1**



Figure D.1: Average total delay in [s]

Table D.1: Average total delay per scenario [s]





#### **D.1.1. Resulting graphs of KPIs**

Figure D.2: Spillback duration in [s]

#### **D.1.2. Base case scenario**

Table D.2: Overview KPIs of basecase







#### **D.1.3. Result X5**

Table D.3: Overview KPIs



#### **D.1.4. Result X10**

Table D.4: Overview KPIs



#### **D.1.5. Result X1000**

Table D.5: Overview KPIs



#### **D.1.6. Result Override**

Table D.6: Overview KPIs





```
Figure D.5: Average total delay in [s]
```
#### **D.2. TFE with bus, experiment demand factor 1.5**

```
Table D.7: Average total delay per scenario [s]
```


#### **D.2.1. Resulting graphs of KPIs**



Figure D.6: Spillback duration in [s]





#### **D.2.2. Base case scenario**

Table D.8: Overview KPIs of basecase



#### **D.2.3. Result X5**

Table D.9: Overview KPIs



#### **D.2.4. Result X10**

Table D.10: Overview KPIs



#### **D.2.5. Result X1000**

Table D.11: Overview KPIs



#### **D.2.6. Result Override**

Table D.12: Overview KPIs



# $\left| \rule{0pt}{10pt} \right.$

## TFE results corridor

#### **E.1. TFE demand factor 1**

#### **E.1.1. Base case scenario**

Table E.1: Overview KPIs of basecase











Table E.2: Basecase delay per intersection [s]

#### **E.1.2. Result X10**

Table E.3: Overview KPIs



Table E.4: Delays per intersection [s]



#### **E.1.3. Result X1000**

Table E.5: Overview KPIs



Table E.6: Delays per intersection [s]



#### **E.1.4. Result override**

Table E.7: Overview KPIs



Table E.8: Delays per intersection [s]







### **E.2. TFE demand 1.0 corridor priority**

Table E.9: Overview KPIs of basecase



Table E.10: Basecase delay per intersection [s]



#### **E.2.1. Result X10**

Table E.11: Overview KPIs



Table E.12: Delays per intersection [s]



#### **E.2.2. Result X1000**

Table E.13: Overview KPIs



Table E.14: Delays per intersection [s]

	<b>Mean</b>	Std.dev	t-value	dF	Significant?
0. 3rd_channel	22.184	4.024	5.005	8.904	Yes: $(p < 0.05)$
1. 3rd_missionrock	15.575	3.417	$-1.953$	10.141	(p < 0.1)
2. 3rd_missionbaynorth	13.885	2.797	2.615	8.318	Yes: $(p < 0.05)$
3. 3rd_missionbaysouth	21.639	3.779	2.131	4.714	(p < 0.1)
4. 3rd_warriorsway	18.996	3.316	0.973	6.720	No
5. 3rd_campus	36.359	6.461	$-3.903$	7.825	Yes: $(p < 0.05)$
6. 3rd_16th	53.572	10.138	$-0.659$	8.682	No
7. 3rd_mariposa	59.136	7.575	$-3.440$	3.100	Yes: $(p < 0.05)$
8. 3rd_18th	39.313	9.758	$-2.126$	6.112	(p < 0.1)
9. 3rd_19th	26.839	3.577	$-0.137$	5.127	No
10. 3rd_20th	31.120	4.061	$-7.501$	9.294	Yes: $(p < 0.05)$
0. 3rd_channel	46.647	4.380	5.073	8.504	Yes: $(p < 0.05)$
1. 3rd_missionrock	38.125	4.075	$-6.205$	8.513	Yes: $(p < 0.05)$
2. 3rd_missionbaynorth	25.530	3.531	1.094	10.186	No
3. 3rd_missionbaysouth	39.633	4.485	$-1.416$	9.442	(p < 0.25)
4. 3rd_warriorsway	40.609	2.506	$-6.142$	4.599	Yes: $(p < 0.05)$
5. 3rd_campus	69.456	4.547	$-3.205$	2.916	(p < 0.1)
6. 3rd_16th	120.333	10.119	$-7.904$	8.831	Yes: $(p < 0.05)$
7. 3rd_mariposa	118.873	7.830	1.508	2.112	No
8. 3rd_18th	75.965	5.909	0.488	3.045	No
9. 3rd_19th	49.512	9.188	$-1.715$	7.794	(p < 0.25)
10. 3rd_20th	52.741	5.985	1.651	6.540	(p < 0.25)
0. 3rd_channel	25.847	3.944	$-2.705$	9.803	Yes: $(p < 0.05)$
1. 3rd_missionrock	19.351	3.635	$-1.244$	6.473	No
2. 3rd_missionbaynorth	15.403	3.830	$-0.026$	8.944	No
3. 3rd_missionbaysouth	22.412	4.160	$-2.323$	5.312	(p < 0.1)
4. 3rd_warriorsway	17.763	4.339	$-2.423$	6.101	(p < 0.1)
5. 3rd_campus	35.534	6.447	$-2.462$	3.495	(p < 0.1)
6. 3rd_16th	63.022	11.970	$-3.380$	6.025	Yes: $(p < 0.05)$
7. 3rd_mariposa	70.388	9.452	$-1.277$	2.986	No
8. 3rd_18th	40.971	9.275	$-2.139$	2.781	(p < 0.25)
9. 3rd_19th	27.634	5.343	$-5.248$	6.433	Yes: $(p < 0.05)$
10. 3rd_20th	31.204	10.924	$-0.457$	8.584	No

#### **E.2.3. Result override**

Table E.15: Overview KPIs



Table E.16: Delays per intersection [s]

	<b>Mean</b>	Std.dev	t-value	dF	Significant?
0. 3rd_channel	22.320	4.332	4.528	8.923	Yes: $(p < 0.05)$
1. 3rd_missionrock	15.572	4.401	$-1.631$	9.489	(p < 0.25)
2. 3rd_missionbaynorth	13.520	2.981	3.359	8.377	Yes: $(p < 0.05)$
3. 3rd_missionbaysouth	22.453	5.758	0.704	6.865	No
4. 3rd_warriorsway	19.044	4.019	0.807	7.338	<b>No</b>
5. 3rd_campus	35.691	5.974	$-3.227$	7.519	Yes: $(p < 0.05)$
6. 3rd_16th	48.435	4.707	6.359	5.880	Yes: $(p < 0.05)$
7. 3rd_mariposa	55.436	6.878	$-0.458$	2.547	No
8. 3rd_18th	38.423	5.920	$-1.797$	2.678	(p < 0.25)
9. 3rd_19th	25.586	4.093	1.858	5.687	(p < 0.25)
10. 3rd_20th	27.004	4.656	0.640	9.187	<b>No</b>
0. 3rd_channel	43.731	4.982	9.455	8.664	Yes: $(p < 0.05)$
1. 3rd_missionrock	36.304	4.433	$-2.657$	8.615	Yes: $(p < 0.05)$
2. 3rd_missionbaynorth	26.396	3.454	$-0.957$	10.276	No
3. 3rd_missionbaysouth	40.713	3.630	$-4.066$	9.908	Yes: $(p < 0.05)$
4. 3rd_warriorsway	38.259	5.083	$-1.277$	7.062	(p < 0.25)
5. 3rd_campus	66.087	5.479	0.778	3.787	No
6. 3rd_16th	119.498	19.232	$-4.008$	8.987	Yes: $(p < 0.05)$
7. 3rd_mariposa	110.852	8.758	6.978	2.726	Yes: $(p < 0.05)$
8. 3rd_18th	78.297	8.456	$-1.572$	5.501	(p < 0.25)
9. 3rd_19th	54.907	8.158	$-7.093$	7.251	Yes: $(p < 0.05)$
10. 3rd_20th	55.869	7.861	$-1.869$	7.864	(p < 0.1)
0. 3rd_channel	23.972	5.337	1.072	9.261	No
1. 3rd_missionrock	16.296	2.820	4.661	5.729	Yes: $(p < 0.05)$
2. 3rd_missionbaynorth	14.920	4.326	0.874	8.961	No
3. 3rd_missionbaysouth	20.610	4.989	0.406	6.254	<b>No</b>
4. 3rd_warriorsway	17.442	4.597	$-1.871$	6.383	(p < 0.25)
5. 3rd_campus	34.662	8.488	$-1.434$	5.476	(p < 0.25)
6. 3rd_16th	57.160	8.390	0.373	3.252	No
7. 3rd_mariposa	60.896	10.764	4.692	3.953	Yes: $(p < 0.05)$
8. 3rd_18th	37.523	8.551	0.129	2.301	No
9. 3rd_19th	24.930	8.047	$-1.241$	8.207	(p < 0.25)
10. 3rd_20th	30.710	7.978	$-0.071$	7.749	No






# **E.3. TFE demand 1.0 with additional bus mode**

#### **E.3.1. Base case scenario**

Table E.17: Overview KPIs of basecase



Table E.18: Base case delay per intersection [s]



#### **E.3.2. Result X10**

Table E.19: Overview KPIs



Table E.20: Delays per intersection [s]



### **E.3.3. Result X1000**

Table E.21: Overview KPIs



Table E.22: Delays per intersection [s]



## **E.3.4. Result override**

Table E.23: Overview KPIs



Table E.24: Delays per intersection [s]

