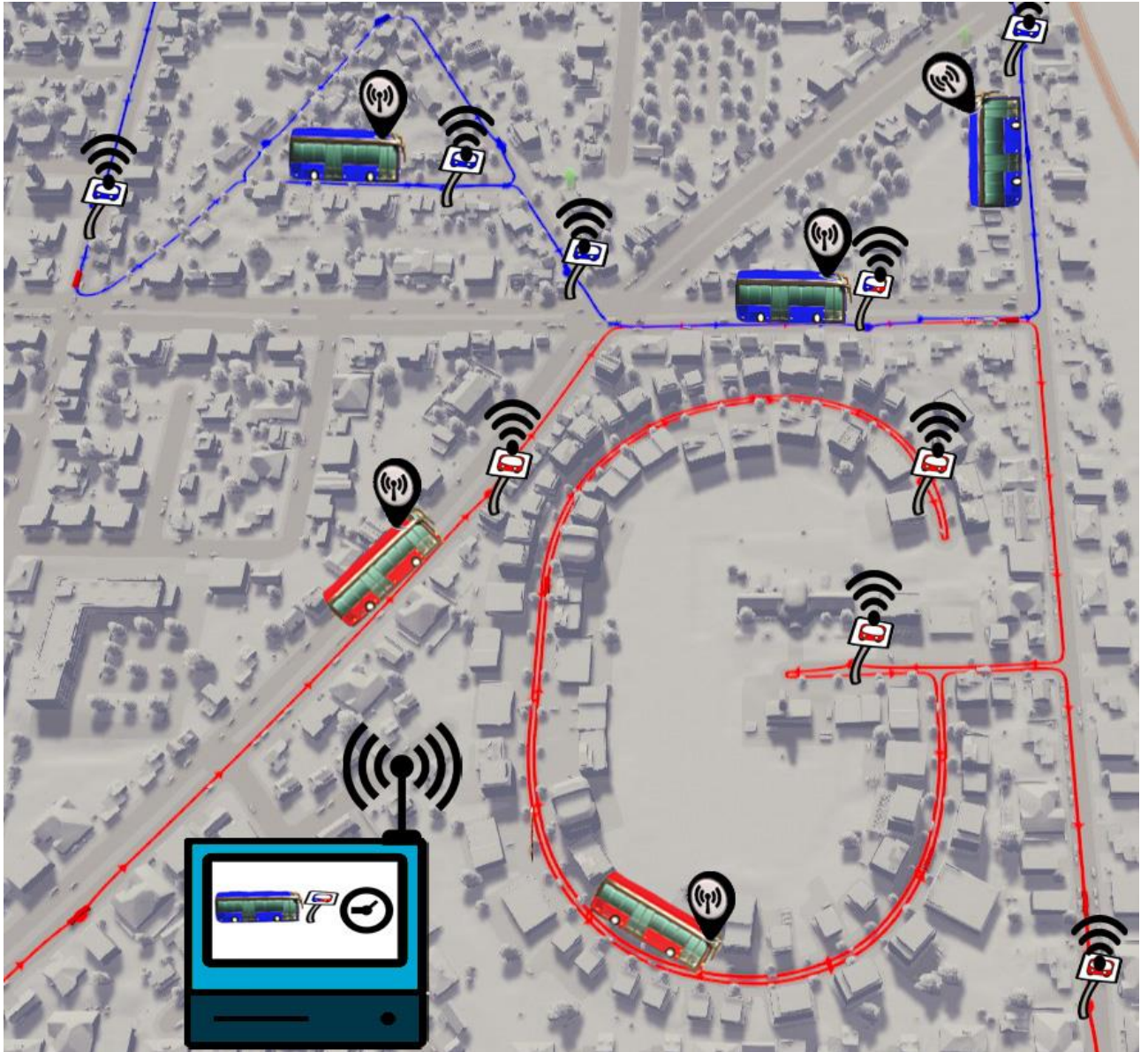# Real-time transfer synchronisation for public transport services using passenger data

A. Gavriilidou

Master Thesis report

August 2016

Real-time transfer synchronisation
for public transport services
using passenger data

Thesis submitted in partial fulfilment of the regulations for the degree of Master of Science

At the Delft University of Technology,

Faculty of Civil Engineering and Geosciences,

Department Transport and Planning.

Student:

Alexandra Gavriilidou (4410718)

Graduation Committee:

Prof. Dr. Ir. Serge Hoogendoorn

Dr. Oded Cats

Dr. Ir. Francesco Corman

Ir. Paul Wiggenraad

Examination Date:

29 August 2016

Cover photo made by the author using the *Cities: Skylines* game (Colossal Order, 2015).

TUDelft

Preface

The idea for this project was conceived during a conference session in 2012, when I was first introduced to the concept of holding a vehicle at a stop until passengers from a delayed connecting service arrive, so that they can transfer without waiting. The discussion was about ways to integrate multi-modal traveller information in the decision process of this holding time. Ever since, I have been wondering about the involvement and influence that passengers can have on this decision in real-time. This motivated the present study, which was shaped such that it would be able to fit within the scope of a Master thesis, yet it gave me valuable insights. If you choose to read further than this page, I hope that you will also find it valuable, and, regardless of the choice you make, I want to thank you for your interest in my research.

The execution of this project was fun and challenging. It had its ups and downs and I am happy to see it's coming to an end. Before this happens, however, credit should be given to everyone involved, directly or indirectly.

The first person I would like to thank is Rob van Nes, who has been a mentor to me since the very beginning of this Master programme. I have made it through the last two years thanks to him and words are not enough to describe how grateful I am to him for all the discussions we have had, for his advice and guidance, for his time and tolerance.

Next, I would like to express my gratitude to my daily supervisor, Oded Cats, for helping me shape and execute this project. Without his ideas, inputs and resources, this project would not have been feasible. Also, his constant support and enthusiasm regarding my progress were catalytic in realising it. In terms of resources, he provided me with the BusMezzo simulation model and brought me in contact with David Leffler, to whom I owe the coupling of my Matlab code to BusMezzo. Thank you David, for making it happen in such a short time and with such a high quality. I really appreciate your time and effort.

I used to believe that thanking the members of the graduation committee is a formality, but having experienced it for myself, I know it is of true essence. The most important contribution of Professor Serge Hoogendoorn was that he believed in me and showed me the way out of a dead-end I had reached. Moreover, his feedback on this report added to the overall quality. During our meetings, he and Oded, along with Francesco Corman and Paul Wiggenraad, posed interesting questions and made valuable comments, which helped me advance with my research, become more thorough and gain insight into more perspectives.

Last but not least, I would like to thank my family and friends for bearing with me throughout the year, hearing me out and brainstorming with me, but also for making me forget about work and just relax.

Alexandra Gavriilidou

Delft, August 2016

Summary

The public transport sector investigates ways to exploit the advancements made in information and communication technologies to improve the performance of the transit system. One such way is the deployment of real-time control strategies that increase the adaptability of the system towards prevailing conditions. Among them, holding strategies are the most frequently studied control method and the focus of the present study. They are used in order to improve on-time performance, eliminate bunching, respond to unexpected demand and prevent passengers from missing their connection.

In real-time holding control strategies, a shift has been noted from vehicle-based strategies, where the vehicle headway regularity is leading, to passenger-based strategies, whose goal is to find a holding time that is the most beneficial for the passengers. Their operationalisation depends on the predictions regarding passenger flows in the network, which may be based on historic or real-time passenger data, the latter being vehicle occupancy or passenger arrivals at stops. However, the influence of real-time passenger data on the performance of transfer synchronisation control has not been investigated. Most previous studies use historic data on passenger flows, while only a few make use of real-time data depending on the measuring equipment assumed available.

Since public transport systems are increasingly equipped with different types of passenger data that can be transmitted in real-time, decisions concerning their deployment and availability should be based on scientific results. To the best of the author's knowledge, no such study has been conducted, where the attainable benefit from the different types of data is investigated. Moreover, there is no controller which considers the riding comfort in the generalised passenger cost function that is being minimised.

The objective of this study is, thus, to develop a holding controller that: (i) considers transfer synchronisation; (ii) makes use of different real-time passenger data sources; (iii) is applicable in real-time, and; (iv) includes the on-board crowding component. This controller is then applied in order to find out which source is the most valuable to acquire in real-time.

The main research question of this study is:

- *What is the effect on the performance of an urban transit network, when different types of real-time passenger data are used in its holding control strategy?*

The development of the controller starts with the definition of the underlying assumptions, which lead to the formulation of the passenger flow prediction scheme and the control decision rules. Some of the assumptions are then lifted in order to increase the realism and functionality of the controller, resulting in four different controller variants. The first one ignores vehicle capacity constraints, which are added in the second. On top of that, the third one takes into account the effect of on-board crowding on the passenger riding experience. The last variant
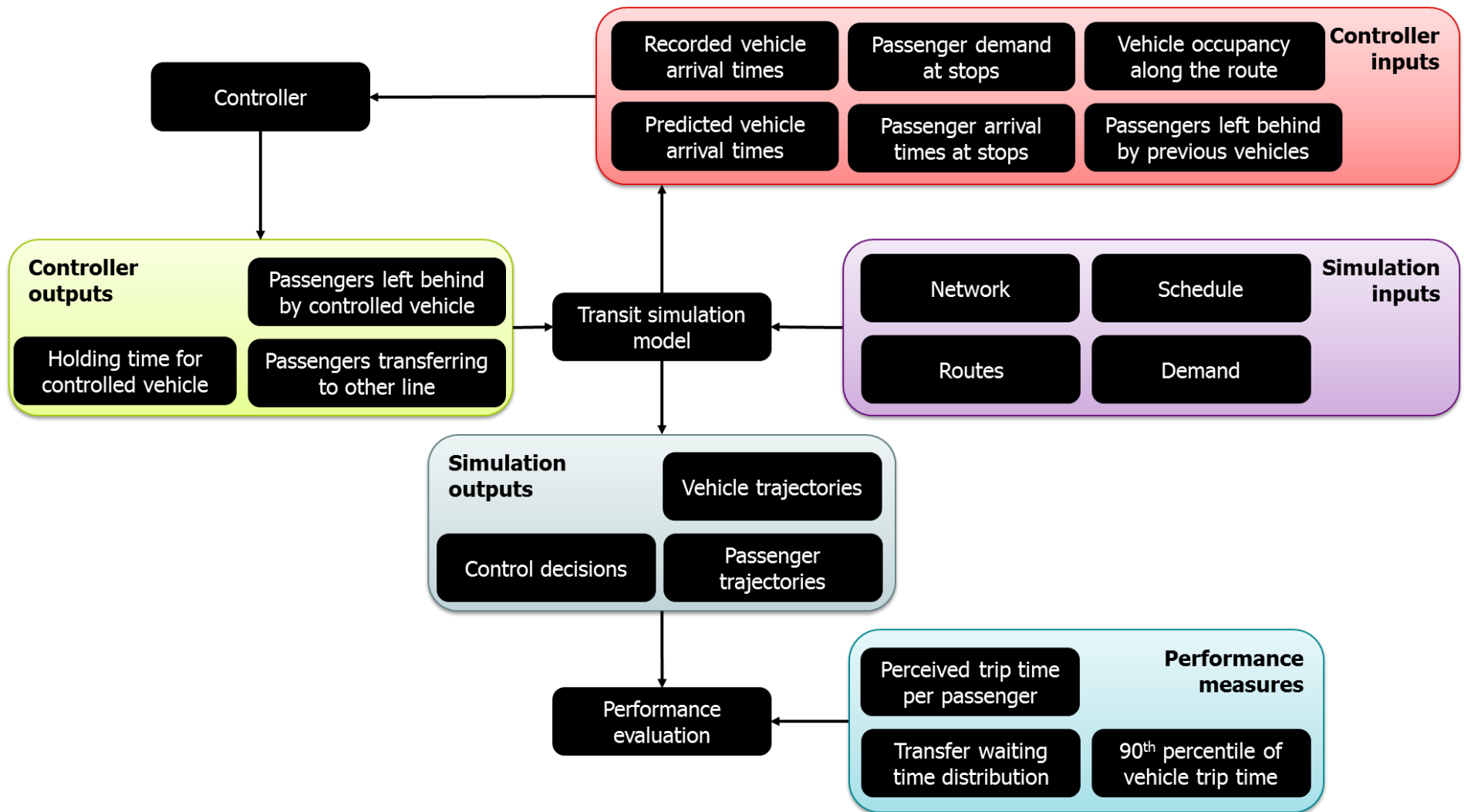
introduces uncertainty in the passenger demand levels by drawing values from a distribution instead of using the historic averages.

Each of these controllers performs two functions. The first function is using the input real-time data to improve the prediction with respect to the passenger flows. Four cases are distinguished regarding the available passenger real-time data types, namely (1) no real-time passenger data, (2) real-time vehicle occupancy, (3) real-time passenger tap-ins, (4) combination of (2) and (3). In all cases, it is assumed that data concerning historical passenger demand is available. The prediction method used is based on these historic estimates, which are replaced by the actual values up to the point that real-time data is available. Using this approach, the prediction is only made for a smaller time horizon, which decreases its uncertainty.

The second function is the execution of a rule-based control algorithm that decides the dispatch time of the vehicle that triggered the controller. The control rules lead to the selection of a holding strategy to ensure either regularity within the line or the synchronisation of a transfer between lines. A discrete set of candidate holding times is defined to represent each of these options. The controller compares the effect each of the candidates has on the passengers and selects the one that has the lowest total generalised passenger travel times. The passenger streams, on whom the effect is estimated, are the passengers on-board, the passengers waiting at downstream stops and the passengers that intend to transfer from the connecting line. It is in this part of the control process that the reliability of the predictions regarding the passengers becomes valuable, since the more reliable the predictions are, the better this effect can be quantified.

The performance of these controllers is assessed by simulating transit operations. The BusMezzo simulation model is used as a testbed for mimicking real-world operations. Each time a transit vehicle enters a transfer stop, BusMezzo calls the controller. The data exchange between them is presented schematically below. The simulation model provides the controller with real-time vehicle and passenger data as well as predictions regarding vehicle arrival times at downstream stops within a predetermined horizon. These outputs are fed into the controller where the passenger data exchange depends on the assumed data availability.

Following its execution, the controller informs the simulation model whether it should hold for regularity or synchronisation and the respective expected vehicle dispatching time. In case of transfer synchronisation, the vehicle is instructed to wait until another vehicle (from either line) arrives at the stop to ensure that the direct transfer will take place, due to the uncertainty associated with the predicted arrival time of the connecting vehicle. The controller stores some of its estimates in order to retrieve them later to gain efficiency.

Controller

Controller inputs
- Recorded vehicle arrival times
- Passenger demand at stops
- Vehicle occupancy along the route
- Predicted vehicle arrival times
- Passenger arrival times at stops
- Passengers left behind by previous vehicles

Controller outputs
- Passengers left behind by controlled vehicle
- Holding time for controlled vehicle
- Passengers transferring to other line

Transit simulation model

Simulation inputs
- Network
- Schedule
- Routes
- Demand

Simulation outputs
- Vehicle trajectories
- Control decisions
- Passenger trajectories

Performance evaluation

Performance measures
- Perceived trip time per passenger
- Transfer waiting time distribution
- 90$^{\text{th}}$ percentile of vehicle trip time

III

At the end of a simulation run, BusMezzo generates output files that summarise the activity of the vehicles and the passengers in the network throughout the simulation. A number of runs are necessary to account for the stochasticity in the simulation model and they are all processed to evaluate the performance of the developed controllers.

Since the objective of the proposed controllers relates to the passenger travelling experience (i.e. minimising their perceived travel time), and especially looks into the synchronisation of transfer, the key performance indicators which are deemed most appropriate concern the passenger activity. These are the following:

a) The perceived trip time from origin to destination per passenger. It is evaluated separately for four passenger groups, since the effect of the control strategy differs among them. The groups use the transfer stop as reference and correspond to passengers (1) alighting upstream, (2) generated downstream, (3) transferring and (4) traversing the transfer stop.

b) The distribution of the transfer waiting time. It checks to what extend the envisaged reduction is achieved.

c) The distribution of the unexpected holding time for synchronisation. This is an indicator of the vehicle arrival time prediction reliability.

d) The 90th percentile of the vehicle trip time per line. It assesses the impact on vehicle operations and is commonly used in determining the fleet size and might be a decisive property for public transport authorities and operators regarding the actual implementation of the strategy.

The next step is the application of the developed controllers on a case study. The selected network examines two tram lines in The Hague, the Netherlands, with a single transfer stop under three demand levels. Each of them is simulated using all combinations of the four controller variants and the four levels of passenger real-time data. A benchmark case is also implemented, where only regularity control is applied at the transfer stop.

The figure below shows the perceived trip time per passenger and passenger group compared to the value obtained in the respective benchmark case when the third controller is used. Negative values correspond to shorter travel times achieved by using the controller, i.e. time savings. The results are presented in order of increasing demand level from left to right. The rows in each part correspond to a passenger group. The results for the upstream group are not shown, since these passengers are not affected by the control strategy that is applied at the transfer stop. The bars within each plot refer to the three passenger data levels and are composed of six time components. These are evaluated for all passengers served within the simulation time from their origin to their destination.

The results show that transferring passengers save on average 2-10 minutes thanks to the proposed strategy, while on-board passengers experience a delay of 1-2 minutes in most

cases. The highest time saving per transferring passenger is obtained when the demand level is low and the controller opts for synchronising more frequently. However, when the demand rises, the results demonstrate that the highest share of synchronisation decisions is not the one that leads to the most savings for the transferring passengers.
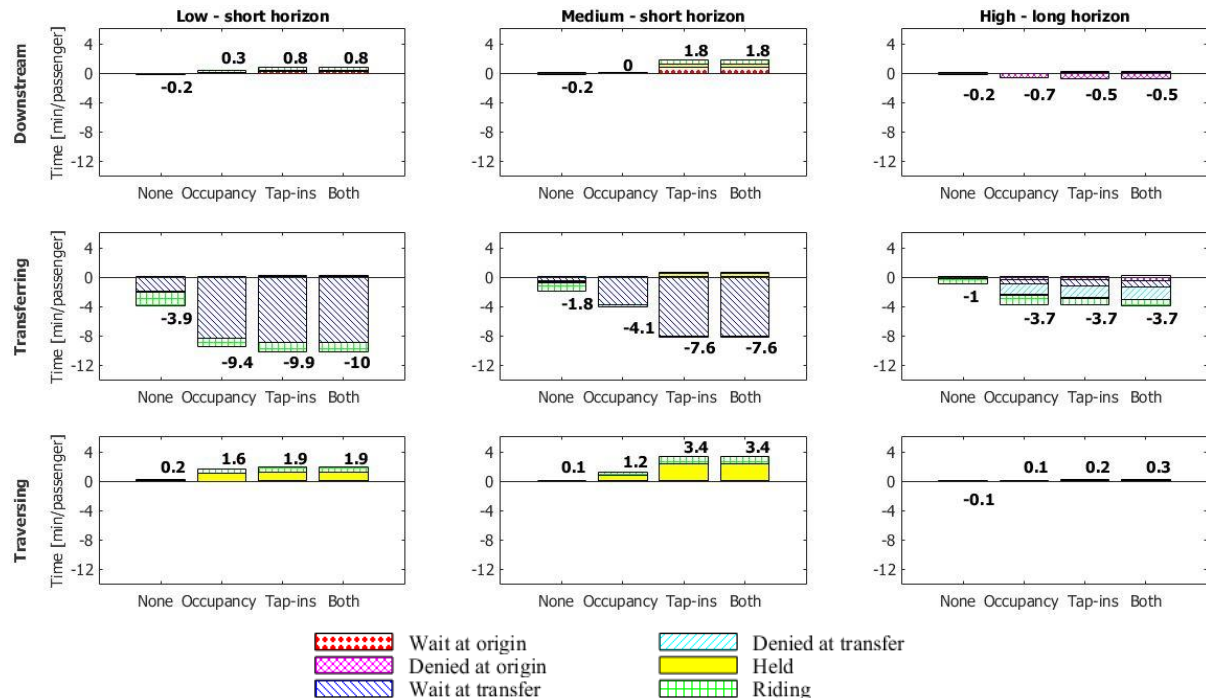
In high network loads, the line operation becomes more irregular and the real-time control of vehicles should focus more on re-establishing regularity rather than looking into the cooperation with other lines. Because of the assumption made by the controllers that the vehicle arrival time predictions are fully reliable, this irregularity has been ignored. The consequence of that is the underestimation of the expected holding time for synchronisation, which leads the controllers to regrettable synchronisation decisions and in some cases creates vehicle bunching.

The comparison of the results for the different passenger data types provides an answer to the main research question. It reveals that the vehicle occupancy is the most valuable real-time passenger data source and it is best exploited when the on-board crowding is considered. Knowing the occupancy of vehicles in real-time can lead to better predictions of the passenger comfort in each control scenario and, hence, better decisions. As the demand rises, the availability of tap-ins leads to more control decisions in favour of synchronisation, which results in a poor controller performance (each downstream passenger needs to wait two more minutes, while traversing passengers are held up to 5min each). This is attributed to the short horizon (10 stops downstream of the transfer stop) adopted by the controllers, which creates a myopic view of the network load. Extending the horizon up to the end of the line expands the view of the load over the network and results in more informed and, therefore, improved

control decisions (almost 4min and 1min saved per transferring and downstream passenger, respectively).

With respect to the four controller variants, the first two controllers perform similarly, except for in the case that the demand is high and passengers may be denied boarding. In this case, the controller with the capacity considerations is more conservative in deciding to synchronise which yields a better performance, i.e. one additional minute saved for each passenger transferring and waiting downstream. The inclusion of on-board crowding influences the controller performance at all demand levels, since over a certain vehicle load the passengers are more comfortable waiting for the next vehicle rather than standing on-board. The introduction of uncertainty in the passenger flow predictions by the fourth controller does not affect the performance, which is attributed either to the meshing resolution or the fact that the same degree of uncertainty has been applied to every passenger flow prediction, thereby neutralising the overall effect.

The main scientific contribution of this study is the inclusion of the passenger level of comfort in the generalised cost function of the controller, which achieves the evening out of the transferring passenger demand among the vehicle fleet. Moreover, the controller can make use of different real-time passenger data sources and adjust its passenger flow prediction model based on the data that is available.

In terms of practical contributions, the developed controller has shown the value of considering the synchronisation of transfers for lines whose schedule is designed such that the vehicle arrivals are synchronised. By applying it, the transferring passengers can save up to a full headway of waiting time at the transfers stop, while only a small delay is experienced by passengers held at the transfer stop. Another contribution is that the controller is applicable in real-time, since its running time is fractions of a second. Last but not least, the conclusion that the vehicle occupancy is the most valuable source to acquire in real-time is valuable for public transport authorities and operators.

Given these contributions, it can be stated that the research objectives of this study have been successfully met and recommendations can be made to the public transport authorities and operators. The first suggestion is to use a controller that takes into account the riding time component in the generalised passenger cost function. This component is based on the on-board crowding conditions, estimates the comfort level of passengers on-board and makes decisions that improve the level of service and spread the transferring passenger demand evenly across the fleet. Given the increased running time of the fourth controller, when the uncertainty in the passenger flow predictions is included, the third controller variant developed in this study is the one that is proposed to be used in practice. An important parameter to be set in the selected controller is the horizon length. When the demand level is expected to be high, a long horizon should be chosen in order to prevent a myopic view of the network load.

Regarding the type of data whose collection in real-time should be prioritised, the study points towards vehicle occupancy. This is nowadays available through automatic passenger counters (APCs), equipment placed on-board vehicles. However, they would have to be adjusted to provide measurements in real-time and the operators should choose the lines they wish to coordinate and equip all their vehicles running during the coordination period with APCs.

When reflecting on the choices made in the controller development, the most important limitation is found to be the negligence of the uncertainty in the vehicle arrival time predictions, which leads to the underestimation of the expected holding time and decisions that further depreciate the service reliability. Moreover, since the prediction uncertainty increases with its spatial and temporal horizon, assigning the same weight to all of them is another choice that would have to be revisited.

Apart from treating these limitations, future research could focus on the development of more advanced controllers on the basis of these ones. The first expansion could be to allow the re-evaluation of a synchronisation decision in case that the originally expected holding time is insufficient. Instead of a binary choice among the candidates, a cut-off point could be estimated, defining the boundary for favouring a transfer synchronisation decision. Additionally, further research should be conducted in order to conclude on the passenger composition at a transfer stop that justifies the implementation of the transfer synchronisation strategy.

An interesting direction for future research would be the inclusion of a real-time passenger data source that has not been accounted for in the present study, namely the actual destination of passengers. This could be obtained by a fare collection system that requires the specification of the full itinerary in advance. Alternatively, their planned choice of route could be crowdsourced, necessitating the consideration of adherence to the original plan and representativeness of the sample, which could, nonetheless, prove to be valuable as the complexity of the assumed network increases.

In terms of practical considerations, it is worth investigating ways to coordinate multiple lines at one stop or treat more transfer stops within one line or even combine the two in a network that features a common corridor. The value of the different real-time passenger data sources should then be estimated again to validate the generalisability of the conclusions made in the present study.

Table of Contents

List of Figures

List of Tables

Notations

R: set of routes

S: set of stops

M: set of vehicles

$\lambda_{s,r}$: passenger arrival rate at stop s in route r

$\rho al_{s,r}$: alighting fraction of on − board passengers at stop s in route r

$\rho tr_{s,i\rightarrow j}$: transferring fraction of passengers alighting at stop s in route i towards route j

$\tau tr_{s,i\rightarrow j}$: passenger transferring time at stop s from route i towards route j

α: allowable headway threshold ratio

$\eta_{i,j}$: desired headway between vehicles i and j

β: generalised cost function coefficients

μ: number of downstream stops in the horizon length

κ_m: capacity of vehicle m

φ_m: seat capacity of vehicle m

γ: on − board crowding multipliers

tnow: current simulation time

thold: holding time of controlled vehicle at the transfer stop

$tarr_{m,s,r}$: arrival time of vehicle m at stop s in route r

$tdw_{m,s,r}$: dwell time of vehicle m at stop s in route r

$tdep_{m,s,r}$: departure time of vehicle m at stop s in route r

$tdr_{i \to j,r}$: running time between stops i and j in route r

$h_{i,j}$: headway between vehicles i and j

$pocc_{m,s,r}$: vehicle occupancy of vehicle m at stop s in route r

$pseat_{m,s,r}$: passengers sitting in vehicle m at stop s in route r

$pstand_{m,s,r}$: passengers standing in vehicle m at stop s in route r

$pdem_{m,s,r}$: passenger demand for vehicle m at stop s in route r

$pbd_{m,s,r}$: passengers boarding vehicle m at stop s in route r

$pal_{m,s,r}$: passengers alighting vehicle m at stop s in route r

$plb_{m,s,r}$: passengers left behind by vehicle m at stop s in route r due to capacity constraints

$ptr_{m,s,i \to j}$: passengers transferring from vehicle m at stop s in route i to route j

$ptrf$: tranferring passengers who fail to board the first connecting vehicle due to capacity constraints

x_{reg}: any variable x when the holding time aims to regularise the service

x_{syn}: any variable x when the holding time aims to synchronise transfers

$\hat{p}occ_{m,s,r}$: real $-$ time vehicle occupancy of vehicle m at stop s in route r

$\hat{p}dem_{m,s,r}$: real $-$ time passenger demand for vehicle m at stop s in route r

$\hat{p}arr_i$: recorded arrival time of passenger i

$Cost_y$: Generalised passenger cost if holding strategy $y \in \{syn, reg\}$ is applied

$TrWait_y$: Transfer waiting time for first vehicle if holding strategy $y \in \{syn, reg\}$ is applied

$TrDenied_y$: Transfer waiting time after boarding denial by first vehicle if holding strategy y $\in \{syn, reg\}$ is applied

$DsWait_y$: Downstream waiting time for first vehicle if holding strategy $y \in \{syn, reg\}$ is applied

$DsDenied_y$: Downstream waiting time after boarding denial by first vehicle if holding strategy y $\in \{syn, reg\}$ is applied

$Held_y$: On $-$ board held time if holding strategy $y \in \{syn, reg\}$ is applied

IvT_{m_y}: Riding time spent by passengers inside vehicle m if holding strategy y $\in \{syn, reg\}$ is applied

1 Introduction

This chapter is an introduction to the present study, starting with the definition of the research scope, objectives and research questions in section 1.1. Section 1.2 describes the approach that is followed to meet these objectives and answer the research questions. The chapter closes by presenting the structure of the report in section 1.3.

1.1 Research scope

In urban environments, there is a growing need to alleviate issues caused by motorised traffic, such as noise and air pollution, congestion and conflicts with vulnerable road users. One way to achieve this is by increasing the share of public transport modes in the modal split, which means that their utility needs to be improved to outweigh that of cars. The utility of public transport services is linked, among other factors, with the regularity (or punctuality depending on the headway) of the service and the door-to-door travel time which may correspond to a multi-leg trip, for which the ability to successfully complete direct transfers is crucial. For this reason, public transport authorities and operators have been investigating ways to improve their service level.

At the same time, the advancements in information and communication technologies (ICT) in the transport sector have enabled the provision of data in real-time and the cooperation of services. The former can be used to dynamically control and influence a process, while the latter allows the process to be viewed and treated from a higher level, as part of an integrated system, and, thus, controlled in a coordinated way.

Therefore, the public transport sector is focusing on ways to exploit these new data sources to improve the performance of the system. One such way is the deployment of real-time control strategies that increase the adaptability of the system towards prevailing conditions. The operationalisation of these strategies depends on the type of information and communication that is available in real-time and can be considered as an input. Most of the research so far has assumed only vehicle to infrastructure (V2I) communication and more specifically Automatic Vehicle Location (AVL), which facilitates the traceability of the vehicle location and improves the estimates of vehicle arrival at a stop.

An unexploited source of information in the decision process for a real-time control strategy is that of the public transport users. As Filippi et al. [2013] have argued, passengers should be empowered and included in a real-time information exchange. According to them, passengers could act as sensors, gathering and communicating data that can be used in improving the system. However, in current practice, the flow of real time information is unidirectional, i.e. the passengers are only on the receiving end of information made available by different types of intelligent transport systems (ITS) along the various stages of their trip (pre-trip, off-vehicle, in-vehicle) [Ambrosino et al., 2014]. Their presence in the public

transport network as well as their decisions and reactions to this information are among the uncertainties of the system that need to be treated statistically. Since such data can nowadays be provided to the real-time controllers by sensors located at stops or on-board public transport vehicles, it is worth investigating the potential they hold. A further step could even be that the data is generated from the passengers themselves by additional features in smartphone applications. These could allow them to communicate their travel plans to the transit agency, for example by accepting a suggested route produced by the application [Daganzo and Anderson, 2016].

The problem statement is that the benefits of including different types of real-time passenger data have not been explored in the literature, yet they hold the answer regarding the type of data whose collection in real-time should be prioritised by the public transport operators in order to improve the level of service. Moreover, there is no controller which considers the riding comfort while focusing on minimising the generalised passenger cost. This can be expressed by an on-board crowding component in the cost function.

The objective of this study is, thus, to develop a holding controller that: (i) considers transfer synchronisation; (ii) makes use of different real-time passenger data sources; (iii) is applicable in real-time, and; (iv) includes the on-board crowding component. This controller can then be applied in order to find out which source is the most valuable to acquire in real-time.

Given this objective, the main research question of this study is:

- *What is the effect on the performance of an urban transit network, when different types of real-time passenger data are used in its holding control strategy?*

In order to answer this question, a number of sub-questions can be formulated, since those help to gain supportive knowledge for answering the main questions, unravel core concepts and bring structure in the research. The formulated sub-questions are the following:

- *According to the literature, how are passengers included in the public transport operational control?*
- *According to the literature, how can prediction methods make use of real-time passenger data?*
- *How can operational control strategies be enriched using real-time passenger data?*
- *Which are the most appropriate performance measures for the effect in question?*
- *How does the performance of the developed controller vary under different conditions regarding the passenger demand distribution and service reliability?*

1.2 Research approach

The aim of this section is to describe the approach followed to meet the objective and answer the research question. The research approach comprises three stages, namely a literature study, the controller development and its application.

The execution of a literature study aids the development of the new controller by providing insights into the different strategies that have been applied in real-time controllers. Emphasis is especially placed on holding control strategies, where transfer synchronisation is sought, in order to guide the determination of the control rules for the new controller. The second part of the literature study focuses on the reviewing of performance measures that have been used by researchers in the evaluation of real-time control strategies. With respect to the modelling of the passenger flows, it needs to be adjusted to make use of real-time passenger data. The applied prediction method is the result of another literature study on ways to improve predictions using real-time data.

In the next stage, a simulation-based approach is adopted, whose framework is depicted in Figure 1.1. It consists of three steps, namely the development of the controller, the coupling of the latter to the simulation model and the controller assessment based on predefined performance measures. During the simulation run, whenever a controlled vehicle enters a controlled stop, the simulation model requests the intervention of the controller to determine the control strategy, while providing real-time data as input.

The controller development starts with the definition of the underlying assumptions, which lead to the formulation of the passenger flow prediction scheme and the control decision rules. Some of the assumptions are then lifted in order to increase the realism and functionality of the controller, which generally performs two functions. One is using the input real-time data to improve the prediction with respect to the passenger flows and the other is the execution of the rule-based control algorithm that decides the dispatch time of the vehicle that triggered the controller. The control rules lead to the selection of a holding strategy to either ensure regularity within the line or a transfer synchronisation between lines at a single point where the different types of real-time passenger data are given as input.

The assessment is performed by applying the developed controller on a case study. Based on the results of the simulation runs and the definition of the most appropriate performance measures, the overall performance of the controller is evaluated. However, the strength of the conclusions drawn depends on the settings and parameters of the controller as well as the network conditions. For this reason, multiple scenarios are simulated, enabling the evaluation of the performance of the controller under different conditions with respect to the passenger demand distribution and service reliability.
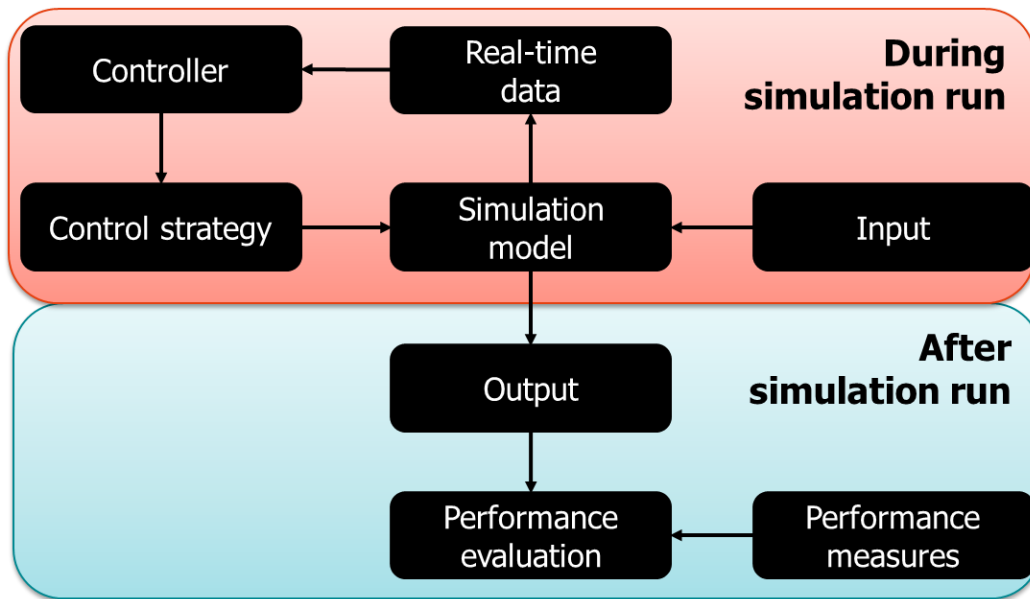
Figure 1.1: Model framework.

The analysis of the simulation results for these scenarios quantifies the effect of using different types of real-time passenger data and answers the main research question. Finally, conclusions are drawn and recommendations are made for the public transport authorities and operators, thereby reaching the objective of the study.

1.3 Thesis structure

The present study is structured as follows. Chapter 2 gives a review of the literature on real-time control strategies and prediction methods. In chapter 3 the controller development and implementation is explained. Chapter 4 describes its application and discusses the results. Finally, chapter 5 summarises the findings of this study, answers the research questions and provides recommendations for the public transport authorities and operators, as well as for future research.

2 Literature review

In this chapter, the literature findings with respect to real-time control strategies and prediction methods are presented. Section 2.1 gives the review on real-time operational control focusing on holding control strategies, which provides useful insights for the controller development and reveals gaps that the present study can fill. In section 2.2, different methods related to the prediction of passenger flows are discussed, which are used as basis for the development of the passenger prediction model. Finally, in section 2.3 the findings of the other two sections are summarised in order to position the present study in relation to the rest of the literature.

2.1 Real-time operational control

As previously mentioned, the deployment of real-time control strategies aims to increase the adaptability of the system towards prevailing conditions. These conditions may reflect deviations from the schedule (timetable) or the creation of an imbalance between supply and demand (overloaded and almost empty vehicles) [Nesheli and Ceder, 2015b].

According to Eberlein [1995], the real-time control strategies can be divided into three categories, namely the station controls, the inter-station controls and others.

1. **Stations controls:** The first category consists of strategies, where the decision variables determine whether specific stops should be visited and what the dispatching time should be. Holding strategies instruct a vehicle to stay for a longer time at a stop, while stop-skipping ones are used to reduce the vehicle trip time by skipping one or more stops of the planned route. In doing so, the waiting time of passengers at the skipped stops is increased and those who wanted to alight there experience the inconvenience of having to alight elsewhere. Deadheading, expressing and short-turning are three forms of station skipping strategies. In deadheading, a vehicle leaves its terminal empty and heads for a designated stop without visiting those in between, while an express vehicle can be dispatched from any stop, and may carry passengers. In short-turning strategies, the vehicle turns around before reaching its terminal, thereby skipping multiple stops in both directions. It is advantageous for passengers in the opposite direction and troublesome for those wanting to board and alight within the skipped segment. Another disadvantage is that it may change the order of vehicles.

2. **Inter-stations controls:** The second category contains strategies, such as speed management and traffic signal control, which affect the travel time of a vehicle between stops. At signalised intersections, the signal phase can be adjusted to prioritise a public transit vehicle that is behind schedule or delay one that is early. In mixed traffic situations, however, the implementation of these methods may be hindered or not yield the desired results.

3. **Others:** The third category includes strategies related to fleet management, such as splitting trains or adding reserve vehicles to cope with disruptions in the system.

Among these strategies, holding ones are the most frequently studied control method and are also favoured by operators over stop-skipping strategies [Manasra, 2015]. They are used to improve on-time performance, eliminate bunching and respond to unexpected demand, while they have an adverse effect on on-board passengers [Ceder, 2007]. Moreover, they are used to prevent passengers from missing their connection, thereby decreasing the transfer time between lines [Ibarra-Rojas et al., 2015]. This, however, may be at the expense not only of on-board passengers but also of those waiting downstream.

There are two types of holding strategies, namely the schedule-based and the headway-based. The former are typical for services with long headways where a timetable is in place and punctuality is important, while the latter is used when the headways are sufficiently short for passengers to arrive randomly at the stops without consulting a schedule [Barnett, 1974]. For urban areas the threshold between short and long headways has been found to be 12min on average, since for larger headways the arrival of passengers at the stop is timed to fit that of the vehicle [O'Flaherty and Mangan, 1970]. For high frequency services, regularity matters more to passengers than punctuality [Abkowitz and Engelstein, 1984].

Several studies have combined holding with other strategies such as stop-skipping [Eberlein et al., 1999], short-turning [Shen and Wilson, 2001], expressing [Sáez et al., 2012], deadheading [Fu et al., 2003] and signal priority [Chandrasekar et al., 2002], and compared the results of their combination with those obtained when each strategy is applied separately. The overall conclusion is that the combined control yields better system performance than any single strategy, yet holding is the single most effective type of intervention [Sánchez-Martínez et al., 2016]. Another combination that led to a similar conclusion is that of holding while applying boarding limits at the stops. The results suggested that such limits should only be implemented in high frequency services and when the next arriving bus is close, in which case they achieve less crowded vehicle loads and reduce the average cycle time and its variability [Delgado et al., 2012].

Due to their superiority and favourable reception, holding strategies are also the focus of the present study. This section discusses the different aspects that are required to implement a holding control strategy. According to Cats [2011], in order to apply a holding strategy, the holding criteria (subsection 2.1.1) as well as the number and the location of the control points (subsection 2.1.2), i.e. the stops where holding can be implemented, need to be determined. Moreover, the types of data that are utilised (subsection 2.1.3) are an important component in their application [Sánchez-Martínez et al., 2016]. The section closes with the presentation of measures that have been used in the literature to evaluate the performance of controllers (subsection 2.1.4).

2.1.1 Holding criteria

Based on the classification by Zolfaghari et al. [2004], holding control strategies can be grouped in two categories with respect to their solution methods. One is referred to as threshold-base control, whereby a set of rules determines the holding time, while the other relies on mathematical programming and optimisation models where the decision variable is the holding time.

The former focuses, in the majority of the studies, on the vehicle movements and aims to achieve punctuality or regularity, depending on the headway type of the service. In schedule-based control, emphasis is placed on adhering to the schedule and vehicles that arrive early at a controlled stop are not allowed to depart prior to their scheduled departure time [van Oort et al., 2012], while in headway-based control the headway between consecutive vehicles is monitored and restored to a desired predefined value by the means of holding. Conventionally, studies consider the headway between the present vehicle and its predecessor [Abkowitz and Lepofsky, 1990; Barnett, 1974; Daganzo, 2009; Fu and Yang, 2002; Turnquist and Blume, 1980], while more recent studies look at the headway between the previous as well as the following vehicle [Cats et al., 2011; Cats et al., 2012; Cortés et al., 2010; Daganzo and Pilachowski, 2011; Guevara and Donoso, 2014; Xuan et al., 2011].

Fu and Yang [2002] found that the optimal headway for the holding strategy lies between 0.6 and 0.8 times the planned headway. Daganzo and Pilachowski [2011] showed that when both the preceding and the subsequent vehicles were taken into account, the controller was able to compensate for large disturbances, such as those due to vehicle breakdowns. Cats et al. [2011] combined the mean headway from the previous and the next bus with that of the planned headway from the previous bus in order to restrict the maximum allowable holding time and found it better than either of the two applied separately. Bartholdi and Eisenstein [2012] proposed a controller that allows the system to express its natural headway which changes dynamically, and repositions the buses in order to equilibrate their headways even in the presence of large service disruptions.

Such control rules can improve the performance of a single line, while in case of multiple connecting lines the rules need to consider the arrival time of the connections. Dessouky et al. [2003] formulated rules to hold a vehicle at a transfer stop so as to synchronise the transfers, i.e. the vehicle is dispatched after the transfers have been successfully completed. Some of the rules considered only the vehicle movements, either their scheduled or their forecasted arrival, while others also took the passengers into account. The latter was performed, in the simplest case, by setting a minimum requirement of transferring passenger volume and, in the more complex ones, by selecting and applying the holding time that would inflict the minimum waiting time for passengers locally or even at downstream stops.

Another approach towards rule-based control with transfer coordination was proposed by Daganzo and Anderson [2016], who used the maximum holding time as a decision variable. They calculate the maximum permissible holding time of a controlled vehicle that arrives at a transfer point and then search for connecting trips within this interval. Only if one or more such trips exist, is the controlled vehicle held and otherwise it is dispatched once its dwell time elapses. In case that a synchronisation decision is made, the vehicle waits for as long as it is required for the transferring passengers to arrive, thereby accounting for an underestimated arrival time prediction.

When the solution method is based on mathematical programming and optimisation, then the objective goes beyond the vehicle movements and checks the effect of a control decision on the passengers. This transition is reasonable, since as the models became more complex and included passenger flows, the controllers could also take them into account, moving from a vehicle-based control towards a passenger-based one.

In such optimisation based controllers, the holding time is the value that minimises an objective function, which may focus only on waiting passengers or balance their interests with those of the on-board passengers [Delgado et al., 2012; Delgado et al., 2009; Eberlein et al., 2001; Sáez et al., 2012] or even consider transferring passengers and their ability to successfully complete a direct transfer [Daganzo and Anderson, 2016; Hadas and Ceder, 2010a; Hall et al., 2001; Manasra, 2015; Yu et al., 2011]. Zolfaghari et al. [2004] were the first to introduce the effect of the vehicle capacity and include in the objective function the extra waiting time of passengers who failed to board the first-arriving vehicle due to the activation of the capacity constraint.

Furth and Muller [2006] proposed the potential waiting time (also called buffer or budget waiting time), whereby the reserved waiting time of passengers arriving at a stop is an extreme (95th percentile), rather than the mean, of the waiting time distribution. Passengers budgeting this additional time minimise the risk of late arrival at their destination. This approach was adopted in the estimation of the passenger waiting time in the holding control strategy developed by Li et al. [2011] for long as well as for short headway services.

These passenger cost components are usually combined in a single objective function in a weighted sum, while the solution horizon can vary from a local perspective, where only the current stop is considered, to a global one, where the impact of the strategy on downstream stops is also taken into consideration. A local perspective was adopted by Zhao et al. [2003] who developed a negotiation algorithm between two agents, one on-board the bus and the other at a stop, who negotiate based on their marginal costs. However, this approach was considered myopic, and they extended it to include a negotiation with passengers at downstream stops as well, but they added smaller weights for them due to the reduced reliability of the estimates further downstream. Delgado et al. [2012] minimised the sum of

the individual travel times of all passengers from the moment they arrive at a stop to the moment they reach their destination during the whole planning horizon, averaged over all involved passengers.

In between the local and global horizons lies the concept of a rolling horizon. It is called 'rolling' because it moves forward as vehicles progress on their routes and 'horizon' due to its finite length [Zolfaghari et al., 2004]. It was adopted by Eberlein et al. [1999] and considered a limited number of vehicles, known as the impact set, in the optimisation, while applying the control strategy only to the first vehicle of the set. Eberlein et al. [2001] developed a deterministic model and, under the assumption of perfect forecast information, found that the effect of holding a vehicle on its consecutive vehicles diminishes quickly and can be sufficiently captured with a very small impact set size, which in their case was three vehicles.

An alternative to the single objective function was implemented by Cortés et al. [2010], who proposed a two-dimensional objective function. The first dimension corresponds to the headway regularity, which affects the passenger waiting time at the stops, while the other minimises the impact on the system, by penalising the extra travel and waiting time inflicted on passengers due to the application of the strategies. Without defining arbitrary weights for each function, they used a genetic algorithm to generate a set of solutions, i.e. a pseudo-optimal Pareto front. The selection among them required the determination of the relative importance of the two objectives.

2.1.2 Control points

The stops along a public transport route where the vehicle dispatch time is subject to regulation, are called control points. Although all stops could be control points, it is required that the vehicle movements are monitored at each one of them, which is not always the case. In order to guide the installation of monitoring and control systems at the stops, the number and location of control points, also referred to as their layout, have been widely studied.

The stops subject to regulation may be a single or a set of predefined stops along the route. In case that holding aims to synchronise transfers, the transfer stop is the control point. Alternatively, they may be decided by the control process when disruptions occur. In practice, the control points are important transfer hubs with high capacity in terms of vehicles [Cats et al., 2011].

Sun and Hickman [2008] showed that it is beneficial in terms of headway regularisation and cost reduction to have multiple control points. That is because, by introducing enough control points to restore the desired headway along the route, there is no need for large corrective actions, which inflict a higher passenger cost [Daganzo, 2009]. Similarly, Cats et al. [2012] found that, if each stop along the route is a control point, then the propagation of discrepancies is prevented by spreading the control over the entire route.

Despite these findings of improved performance when multiple control points are used, Cats et al. [2014] found that the performance is more sensitive to the location of the control points rather than their number. This could be attributed to service characteristics but also to the holding control strategy that was implemented. Higher service uncertainty could require more control points, while a strategy that considered the positions of all the vehicles, instead of just the previous and the following, could render the location of the control points less important.

Regarding the location of the control points, there is general agreement in the literature that the control points should precede a sequence of high-demand stops [Abkowitz and Engelstein, 1984; Liu and Wirasinghe, 2001; Turnquist and Blume, 1980]. Furth and Muller [2008] also concluded that the control points should be at stops with high boarding rates, preferably located at early stops along the route. Last but not least, stops with high through-passenger demand should be excluded, because of the negative effect of holding on on-board passengers [Hickman, 2001].

2.1.3 Utilised information

Prior to the existence of intelligent transport systems (ITS) that can track the position of vehicles (automatic vehicle location systems, AVL) and the amount of passengers (automatic passenger counters, APCs), the application of real-time control strategies required personnel strategically located to make the control decisions [Abkowitz and Lepofsky, 1990]. Most of those early studies assumed that the controller had little or no real-time information on the position of vehicles along the line and the holding strategy was applied at pre-specified control points on the basis of the timetable and possibly the distance between consecutive vehicles [Carrel et al., 2010]. As noted by Bartholdi and Eisenstein [2012], the objective was generally to reduce the variation in the distribution of observed headways.

Once AVL systems became available, studies have assumed accurate and real-time knowledge of the vehicle locations which enables the headway regulation, especially in cases where the headway to the preceding vehicle is considered. Knowledge of the current location can also enhance the prediction of the arrival time of the next vehicle or connecting services which may be used in the controller. Yu and Yang [2009] developed a holding strategy, which holds vehicles arriving ahead of schedule, while ensuring their on-time performance at the next stop. This is done by forecasting the departure time of the controlled vehicle from the next stop using a model based on support vector machine (SVM).

Strategies that aim to minimise any passenger cost function require information about the passengers. This corresponds to vehicle loads and passenger arrival rates, as well as alighting and transferring fractions at each stop. These are conventionally expressed by offline estimates, while the shift from vehicle- to passenger-based control has necessitated the real-time tracking of passenger flows. New technological developments have facilitated the

provision of such real-time data about the passengers. APCs are the most widespread sources, since they are the initial vehicle-based passenger tracking solution, relying on equipment placed on-board, while the emerging availability of smart-card data could mark the transition towards a passenger-based tracking mechanism.

Zhao et al. [2003] used APCs to count the passengers getting on and off at each stop, while they assumed information exchange between stops and buses communicating via a WAN (Wide Area Network). Information, such as the most recent bus departure time and the number of passengers at the stops, is fed into the controller to decide the holding time. When the buses are in between stops, AVI (Automatic Vehicle Identification) or GPS (Global Position System) data is utilised.

Sáez et al. [2012] combined offline historical data with dynamic online data. The attainability of the latter presumed that sensors were placed in the vehicles and at the stops and their measurements were communicated to the controller, which was seen as a central dispatcher.

Dessouky et al. [2003] compared seven schedule-based bus-holding strategies which used different levels of information and concluded that the best strategy was the one which made use of AVL and APCs as real-time data sources and adopted a global perspective. They used these data to forecast the bus arrival times and the passenger demand at the transfer locations and downstream stops. The demand forecast was based on mean passenger arrival rates and predetermined transfer and alighting probabilities which were applied on real-time occupancy data.

2.1.4 Performance measures

The evaluation of developed control strategies is normally performed through simulation. Since the first controllers considered only vehicle movements, the indicators that were initially used, evaluated the first-order effect on the vehicles themselves, and deduced the impact on passengers. However, as noted by Sánchez-Martínez et al. [2016], the indicators that are typically used, measure the effect on passengers (waiting times and trip times) and are often complemented by measures of headway regularity and vehicle loads.

A key performance indicator for headway-based control is headway variation. As Abkowitz and Lepofsky [1990] argued, the mean headway should remain unaffected by the applied control, unless vehicle arrival time lateness propagated severely during the day, causing an overall increase of the average headways. For this reason, its variation is used instead. By reducing the headway variance, the passenger loads among consecutive vehicles become more equally distributed downstream of the control point and passengers experience less crowding [Li et al., 2011]. The coefficient of variation of the headway averaged over all stops is another performance measure, which provides a sound and normalised measure of service

regularity [Fadaei and Cats, 2016]. It is the ratio between the standard deviation and the mean actual headway.

Moreover, the headway adherence, i.e. the share of vehicles that arrive with a headway within a certain percentage of deviation from the planned headway, is a regularity indicator [Cats, 2014]. TCRP [2003] proposes a 50% deviation for the definition of this range. Milkovits and Wilson [2010] used, instead, the share of headways which were shorter than 60 seconds or longer than twice the planned headway. TCRP [2003] defined an ordinal scale for the level of service (LOS) scores in terms of headway regularity, which depends on the coefficient of variation and the share of bunching.

In case of a schedule-based control, an appropriate performance indicator is schedule adherence. This compares the vehicle arrival to the timetable and considers adherence when the arrival lies within a time window of 1 minute early and 3 minutes late of the scheduled arrival time [Cats, 2011].

The effect on passengers may distinguish between groups of passengers (on-board or waiting) or it may consider the overall trip time. Abkowitz and Lepofsky [1990] used the delay time at the control points to measure the delay inflicted on on-board passengers due to holding and the waiting time of passengers at downstream stops. The latter is generally expected to be reduced once the service is regulated, because vehicle bunching is prevented. In cases that passenger flows are not modelled, the computation of the average waiting time is approximated as a function of the service headway and its coefficient of variation [Li et al., 2011].

Dessouky et al. [2003] used the average passenger trip time from origin to destination and the total passenger delay, which is composed of the delays experienced at each trip phase, namely at the originating stop, on-board and transferring to another line. The delay, if any, for originating passengers is the difference between their actual departure time and the scheduled departure time of the next vehicle following their arrival at the stop. If on-board passengers experience delay, that is again the difference between the actual and the scheduled departure time, but in this case the bus lateness at the previous stop is subtracted to prevent double-counting the delay for multiple stops. For transferring passengers it is assumed that the schedules of the connecting lines are coordinated and passengers that intend to transfer plan to board the vehicle that provides this direct transfer. A late arrival of their first vehicle at the transfer location may, then, lead them to miss their planned connection and, thus, the delay is the difference between their actual departure time and the scheduled departure time of the vehicle they were planning to catch. Nesheli and Ceder [2015a] combined these two measures and introduced the average additional travel time per passenger, which is the average extra time experienced by passengers throughout their trip (from origin to destination) due to service irregularity.

With respect to transfers, Hadas and Ceder [2010a] used the number of direct transfers as a performance measure, analysed in terms of the number of transfers and the flow of passengers in direct transfers. Alternatively, the total number of simultaneous arrivals of vehicles of connecting lines at a transfer point has been used [Manasra, 2015]. Another indicator introduced in the controller developed by Daganzo and Anderson [2016], which could also be applied in the evaluation phase is the recovery factor. It reflects the portion of the holding time experienced by the average affected passenger, under the assumption that passengers care only about the deviation of their actual arrival time at their destination from their scheduled one. Consequently, if the vehicle manages to make up for the holding time before each of those affected by the holding alights, then the performance result is positive.

Last but not least, another performance measure is the average standing time per passenger [Cats et al., 2012]. That is used as a crowding measure to capture the level of comfort of passengers. Its computation requires the tracking of the vehicle occupancy and knowledge of the number of seats per vehicle.

2.2 Predictions about passengers

Passenger flows are an integral part of the transport network and, as can be inferred from the review of real-time control strategies (section 2.1), the effect of a control decision on the passengers is either considered in the selection of a control strategy or in the evaluation of its performance. The former is of interest in this section, since it necessitates the prediction of future states, while the latter only handles elapsed events.

The prediction method may differ regarding the level of representation of the passenger flows as well as the type of data that is used. Passengers may be represented on a disaggregate level or they may be aggregated per stop and line, while the data might be historic, real-time or their combination.

Every passenger trip can be characterised by its spatial and temporal origin, its destination and, if applicable, the stations, where transfers were made between lines. When aggregated, the passenger demand can be represented by arrival rates at each stop, possibly extended by additionally considering the destination, hence having an arrival rate per origin-destination (OD) pair. This presumes an OD matrix between each pair of stops for every line and a transfer matrix between lines at the transfer stations. As noted by Sánchez-Martínez et al. [2016], the research done by Gordon et al. [2013] handles the estimation of OD matrices using automatically collected data, while McCord et al. [2010] propose iterative proportional fitting for the matrix estimation when only boarding and alighting data is available. The destination can, alternatively, be specified as a function of the on-board vehicle occupancy, using alighting rates for each stop. Transferring passenger flows can also be represented by a transferring rate being a function of either the vehicle occupancy or the alighting passengers.

In the greatest part of the literature, these rates are derived from offline historic data and appear as deterministic values, static over the studied period. Stochasticity is sometimes incorporated by assuming that these rates follow a certain distribution [Hall et al., 2001; Hickman, 2001; Zhao et al., 2003]. Passenger demand dynamics were considered in the prediction scheme and the control process by Sánchez-Martínez et al. [2016]. They used time-varying mean arrival rates at the OD-level and found that the inclusion of dynamics can improve the performance compared to static inputs only when the dynamics lead to significant overcrowding.

In the absence of real-time data, the prediction of the passenger demand at stops is the product of the passenger arrival rate and the vehicle headway. The number of alighting or transferring passengers depends on the prediction of the vehicle occupancy and the respective rates. The vehicle occupancy is predicted recursively at every stop based on the occupancy at the previous stop and the forecast regarding boarding and alighting passengers. The boarding passenger flow is the passenger demand, restricted by the remaining vehicle capacity, if the latter is taken into consideration.

What is of particular interest in the present study, however, regarding the passenger flow predictions, is the use of real-time data to either update these offline historic estimates or to replace them. Three data types regarding the passengers may be acquired in real-time, namely vehicle occupancy, passenger arrivals and passenger destinations. More specifically, the first type, the vehicle occupancy, represents the number of passengers that are on-board a specific vehicle and can be derived from the APCs [Hwang et al., 2006].

The second is the number of passengers arriving at a stop and it can be retrieved either by the number of 'tap-ins' at a stop or by sensors located at the stop, which monitor the passenger movements around it. The former requires that the tickets are validated upon arrival at the stop instead of inside the vehicle, as is the case, for example, in metro systems, while the latter could, for instance, be a camera whose footage is processed in real-time to compute the number of detected passengers.

The third type corresponds to the destination of the passengers, which calls for a fare collection system that charges in advance for the distance intended to be travelled, so the passenger has to specify the alighting destination. Potential transfers may be inferred thereof or in some cases, mostly observed in long-distance travel, the tickets may also specify the entire itinerary, including the transfer points. Moreover, if a stop serves more than one line, then the generated demand at the stop can be assigned to specific lines when the destination is known.

It is worth noting that even if all these data types are available in real-time, they can only improve the estimate of the current state of the network. Short-term predictions are still needed to compute the effect of candidate holding strategies in the near future. The prediction

approaches are generally divided into two categories, namely the data-driven (sometimes referred to as non-parametric) and the model-based (also known as parametric). As explained by van Lint [2004], the main difference between the two approaches is that the data-driven approaches consider the underlying processes as black boxes and exploit purely inductive techniques for their predictions, while model-based approaches explicitly address the physical mechanism that governs these processes. The methods that have been used in the literature, applying either of the two approaches, are presented in the remainder of this section. More specifically, subsection 2.2.1 refers to data-driven methods, while subsection 2.2.2 to model-based ones.

2.2.1 Data-driven methods

The common characteristic of the data-driven methods is that they correlate mean (observed) values to current and past data by means of statistical inference or artificial intelligence techniques [van Lint, 2004]. Time series analysis focuses on historical data from previous days to estimate the most likely future state, while artificial intelligence methods, such as support vector machines (SVM) and artificial neural networks (ANN), use clustering and regression to self-define relations between empirical variables [Hans et al., 2015]. Yu et al. [2011] state that SVM and ANN appear to be promising approaches to describe complex systems such as transit systems, due to their versatile structures and adaptive learning processes. However, most of the research focuses on vehicle arrival time predictions using time series, ANN, and Kalman filtering techniques [Yu and Yang, 2009], while little attention has been paid in the literature to passenger demand forecasting [Ma et al., 2014; Wei and Chen, 2012].

Sáez et al. [2012] used Autoregressive Integrated Moving Average (ARIMA) models as a method to forecast passenger demand. They developed a hybrid predictive control strategy, where the passenger demand was modelled as a disturbance and whose prediction combined offline historical data with dynamic online data. Their methodology required a good estimation of the OD matrix for each modelling period from historical data. Online data, regarding the boarding and alighting flows as well as the passenger arrival time at stops, were collected from sensors located at stops and buses and used complementarily. With the application of ARIMA models, the offline estimations were corrected based on the observed preferences of the passengers already in the system and used for the real-time passenger demand prediction.

Ma et al. [2014] developed a pattern hybrid approach based on the Interacting Multiple Models (IMM) algorithm. Hybrid refers to the mixture of offline and real-time data, while the approach consists of four steps. In the first step, time series are constructed from historical data for the weekly, daily and hourly demand patterns. The respective pattern models are then created, such that the time series are adequately represented. In the next step historical data and the pattern models are used to calibrate the transition probability matrix, which

expresses the probability that a transition occurs from one model to another at a specific time. In the last step the IMM algorithm generates the prediction by combining the pattern models, the probability transition matrix and real-time data. When combining the patterns models, weights are being used which in the case of the IMM algorithm are dynamically updated. This approach is intended for short-term demand prediction, whose viability for real-time applications depends on the on-line data acquisition time. The developed model assumed intervals of 30min and needs to be extended to handle shorter ones.

Wei and Chen [2012] used a hybrid forecasting approach which combined empirical mode decomposition (EMD) and back-propagation neural networks (BPN) to predict the short-term passenger flow in metro systems. The approach consists of three stages. The first (EMD Stage) decomposes the short-term passenger flow series data into a finite and small number of intrinsic mode function (IMF) components, which express oscillatory modes of the original time series data based on the local characteristic time scale. The second (Component Identification Stage) identifies the meaningful IMFs by using statistical methods to measure the correlation between each IMF component and the original time series data. The meaningful IMFs, defined as those with the highest correlation, are used as input for the application of BPN in the third stage (BPN Stage) which performs the passenger flow forecasting. The results of these forecasts can then be used by the service operators to adjust their operational planning.

2.2.2 Model-based methods

As previously mentioned, model-based approaches focus on the theory that underpins the process. The three processes that characterise the passenger flows, namely their arrival, their volume on-board vehicles and their route choice, are separately discussed in the ensuing.

Passenger arrival

Since the process in question refers to passenger flows, and hence humans, behavioural models from discrete choice theory are a common modelling approach. Utility theory has been used by researchers for the prediction of the passenger arrival time, which is useful in the estimation of the impact of a control strategy on the passenger demand at downstream stops.

Fonzone et al. [2015] proposed a passenger utility function that is based on an anticipated risk-averse waiting time. It is named 'anticipated', because it refers to their perceived service reliability regarding the bus departure time, and 'risk-averse', because passengers aim to minimise the probability of missing the bus due to their own late arrival.

It might be argued that such a model is only applicable for long headway services, since, as already discussed in section 2.1, in this case they consult the schedule and are otherwise assumed to arrive uniformly at the stop. However, a review by Luethi et al. [2007] found that passengers check the schedule, or the real-time information systems, even when the service headway is 5min. The online availability of real-time information for passengers, through for

example journey planners, even before they reach the stops increases their awareness of the network conditions. Therefore, as the diffusion of internet-based information and smartphones increases, it can be expected that the passengers will have knowledge of the bus departure time in advance and will attempt to coordinate their arrival time respectively.

On-board passenger volume

In cases when the alighting and transferring passenger flows are expressed by rates as a function of the vehicle occupancy, it is important to predict the latter in real-time. In a study by Moreira-Matias and Cats [2016], the authors argue that APCs are not placed in all vehicles of the fleet and their records are usually not communicated in real-time since their design and deployment was meant to support tactical planning. In order to overcome the shortcoming of the vehicle load direct measurement, they proposed the use of AVL data for its estimation. The modelling approach constructs the load profile by reverse engineering the dwell times as function of passengers flows through local constrained regression.

Another approach was adopted by Hans et al. [2015]. They make use of a particle filter (PF), a general Monte Carlo (sampling) method, which computes all possible system states, and provides the most likely one. Based on a stochastic bus operation simulation model, the PF randomly generates a set of possible vehicle arrival times, which are called particles. The bus load of traversed stops is then estimated by weighing the contribution of each particle according to the difference between the estimated arrival time and the actual one.

Passenger route choice

The elements that are involved in the passenger route choice are their origin and destination stops, as well as the transit path used in-between them. The most explicit way to represent these choices is using transit assignment models. An overview of these models and their evolution in time can be found in the review by Liu et al. [2010], yet they have not been used in the control context.

The most advanced of these models are the agent-based simulation models, which comprise the strategies of individual agents and the interactions with each other and with their environment. Each agent makes a sequence of en-route travel decisions according to personal preferences as well as expectations that are shaped based on real-time information availability and day-to-day learning processes.

Such disaggregate and adaptive representation of the passengers' choices may, however, contain unnecessary complexity for a predictive scheme that provides input into the real-time controller, which calls for the development of alternative approaches for the predictions.

2.3 Synthesis

The review of holding control strategies leads to the conclusion that early studies focused on the vehicle regulation such that the service operation would be restored within desired thresholds and the solution method towards the determination of the holding time was based on rules. In more recent studies, a shift is observed from vehicle-based control objectives to passenger oriented ones. The impact of a strategy on different passenger streams is explicitly modelled and optimisation techniques are adopted to find the holding time which would benefit them the most.

The passenger streams that the several studies consider depend on their scope and modelling realism. The former may, for example, refer to whether a single or multiple lines are controlled and, consequently, whether transferring passenger flows are included, while the latter could encompass the consideration of vehicle capacity constraints and passengers who are denied boarding. The overall modelling complexity with respect to the passengers varies across the studies and depends on whether stochasticity, dynamic effects and real-time data are taken into account.

Table 2.1 summarises these aspects for the reviewed studies. The first column contains a reference to the respective study. The second and third describe the solution approach, the method that was used and the objective, respectively. The method is categorised into rule-based and optimisation. The former approach follows a set of if-then rules which lead to the control decision, while the latter follows an iterative process which minimises an objective function to reach the control decision. The optimisation approach may be performed online or offline, which is represented in the table by 'Opt − On' and 'Opt − Off' respectively. This classification depends on whether a rolling horizon is being applied or not. The rolling horizon framework comprises the optimisation of the holding time of multiple vehicles within a horizon and the implementation only of the one that corresponds to the vehicle that triggered the controller. The other decisions are discarded and recalculated when another vehicle activates the controller. This approach is referred to as online optimisation.

The objective is grouped into vehicle-based (abbreviated as 'VB') and passenger-based (abbreviated as 'PB'). When the holding strategy is aimed at regularising the service, the focus is placed on the headways of consecutive vehicles and, therefore, the objective is considered vehicle-based. In the case of a passenger-based objective, the goal is to find a holding time that is the most beneficial for the passengers. This overview highlights the strong link between the method and the objective, with the only exceptions being the study by Dessouky et al. [2003] and Daganzo and Anderson [2016], who used a rule-based approach to synchronise transfers, while looking at the effect on passengers.

The rest of the columns in the table describe how passengers are taken into account in the controller of each study. The 'Capacity' column describes whether the controller considered

vehicle capacity constraints or not. The 'Transfers' column refers to the consideration, or not, of transferring flows between lines in the network in the objective function. The 'Uncertainty' column refers to the way the passenger arrival and alighting estimates were computed by the controller, namely if they took uncertainty into account or followed a deterministic approach. In 'Dynamics', the inclusion, or not, in the controller of time-dependent passenger arrival and alighting rates is reflected. The 'Real-time' column shows whether the controller utilises data with respect to the passengers that are available in real-time.

In the cases of vehicle-based ('VB') objectives, the passengers were not part of the controller, so these columns are not filled. It should, however, be noted, that in most of these studies the passengers were modelled in the simulation and the effect of the applied strategy on them was quantified using passenger-related performance measures at the evaluation phase.

It becomes obvious that only very few studies have developed strategies that can make use of real-time data regarding the passengers, while the type of data that each of them uses is shown in Table 2.2. Yet, they do not consider either capacity constraints or transferring passengers and there has been no evaluation of the benefit achieved in the control decision due to the availability and inclusion of each of these types. Moreover, the riding comfort based on on-board crowding conditions has so far only been used in the performance evaluation of control strategies and not in the generalised passenger cost function.

The part of the literature study regarding the prediction methods for passenger flows using real-time data revealed that this research field is still in its early stages. The predictions are mainly focusing on the vehicles and their arrival times, delaying the shift towards passenger-oriented predictions. Most of the studies either use statistical methods to update the historic estimates with the online data or they attempt to make predictions about the passengers by looking at the vehicle movements.

Table 2.1: Passenger modelling in real-time holding controllers.

[Author(s), Year]	Solution method	Objective	Capacity	Transfers	Uncertainty	Dynamics	Real-time
[Abkowitz and Lepofsky, 1990]	Rule-based	VB	-	-	-	-	-
[Bartholdi and Eisenstein, 2012]	Rule-based	VB	-	-	-	-	-
[Cats et al., 2011; Cats et al., 2012]	Rule-based	VB	-	-	-	-	-
[Cortés et al., 2010]	Opt – On	PB	No	No	Yes	Yes	Yes
[Daganzo, 2009]	Rule-based	VB	-	-	-	-	-
[Daganzo and Pilachowski, 2011]	Rule-based	VB	-	-	-	-	-
[Daganzo and Anderson, 2016]	Rule-based	PB	No	Yes	Yes	No	Yes
[Delgado et al., 2012]	Opt – On	PB	Yes	No	No	No	No
[Dessouky et al., 2003]	Rule-based	PB	No	Yes	No	No	Yes
[Eberlein et al., 2001]	Opt – On	PB	No	No	No	No	No
[Fu and Yang, 2002]	Rule-based	VB	-	-	-	-	-
[Guevara and Donoso, 2014]	Rule-based	VB	-	-	-	-	-
[Hadas and Ceder, 2010a]	Opt – Off	PB	No	Yes	No	No	No
[Hall et al., 2001]	Opt – Off	PB	No	Yes	No	No	No
[Hickman, 2001]	Opt – Off	PB	No	No	Yes	No	Yes
[Li et al., 2011]	Opt – Off	PB	No	No	No	No	No
[Liu et al., 2014]	Opt – Off	PB	No	Yes	No	No	No
[Manasra, 2015]	Opt – On	PB	Yes	Yes	No	No	No
[Nesheli and Ceder, 2015b]	Opt – On	PB	Yes	Yes	No	No	No
[Sáez et al., 2012]	Opt – On	PB	Yes	No	Yes	Yes	Yes
[Sánchez-Martínez et al., 2016]	Opt – On	PB	Yes	No	No	Yes	No
[Sun and Hickman, 2008]	Opt – Off	PB	No	No	No	No	No
[van Oort et al., 2012]	Rule-based	VB	-	-	-	-	-
[Xuan et al., 2011]	Rule-based	VB	-	-	-	-	-
[Yu and Yang, 2009]	Opt – Off	PB	Yes	No	No	No	No
[Yu et al., 2011]	Opt – Off	PB	Yes	Yes	No	No	No
[Zhao et al., 2003]	Opt – Off	PB	No	No	Yes	Yes	Yes
[Zolfaghari et al., 2004]	Opt – Off	PB	Yes	No	No	No	No

Table 2.2: Real-time passenger data used in real-time holding controllers.

[Author(s), Year]	Vehicle Occupancy	Passenger demand	Alighting passengers
[Cortés et al., 2010]	x	x	x
[Daganzo and Anderson, 2016]	x		
[Dessouky et al., 2003]	x		
[Hickman, 2001]	x	x	
[Sáez et al., 2012]	x	x	x
[Zhao et al., 2003]	x		

The present research develops a holding controller that aims to synchronise transfers or simply regularise the service, depending on which has the lowest generalised passenger cost. In the latter, the on-board crowding component is also included. Since the decision needs to be made fast, for the controller to be applicable in real-time, it is designed to be rule-based. The controller makes use of different types of real-time data about the passengers, either separately or in combinations, and predicts the evolution of the system in order to determine the most convenient holding time. The prediction method used is based on the historic estimates, which are replaced by the actual values up to the point that real-time data is available. Using this approach, the prediction is only made for a smaller time horizon, which decreases its uncertainty.

3 Controller development and implementation

In this chapter, each of the components of the model framework depicted in Figure 1.1 is discussed in detail. Section 3.1 describes the development of the controller. The simulation model and its coupling to the controller are discussed in section 3.2. In section 3.3, the process to evaluate the new controller is explained.

3.1 Controller development

The new controller aims to make use of different types of real-time data about the passengers in the transport network and select the holding time for the controlled vehicle following a rule-based approach. This decision concerns vehicles that reach a transfer stop and may be held to synchronise transfers, i.e. to offer a direct transfer to passengers from connecting lines. An example of such a situation is depicted in Figure 3.1, where bus routes r_1 and r_2 intersect at the transfer stop s_{12}. The bus m_1, following the first route, is currently located at s_{12} and the controller needs to decide whether to hold it until passengers on-board the m_2 bus of the other route can directly transfer at this stop. This example is used as reference in the ensuing.
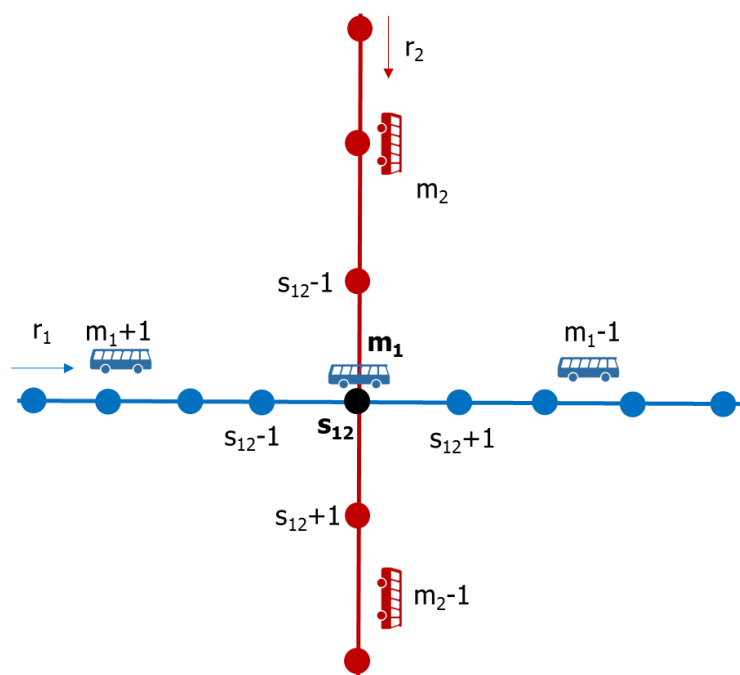
Figure 3.1: Example of a control case of a two-line network with a single transfer stop.

For ease of reference, this controller, which applies a rule-based holding control strategy for transfer synchronisation using real-time passenger data, is hereafter called PassIT. Its description is given in subsection 3.1.1 and it is accompanied by a number of assumptions which are made to simplify the decision process. Some of these assumptions are relaxed and the modifications made to PassIT are presented in an order of increasing complexity in the subsections 3.1.2 – 3.1.4. A comparison of the developed controllers is given in 3.1.5.

3.1.1 PassIT

In this subsection the assumptions of the PassIT controller are first presented. Then, the prediction model with respect to the passenger flows is described, followed by the control rules that govern the decision process towards a holding control strategy.

Assumptions

To begin with, it is assumed that AVL data is available for the entire fleet and as a result, the vehicle location, arrival and departure times at visited stops are available in real-time and are assumed to be reliable.

The passenger arrival process is assumed to be random, implying that the service is frequent and the availability of real-time information systems does not affect the passenger arrival process. Passengers are represented by arrival, alighting and transferring rates per stop and route, which are assumed to be known and fixed over the study period. When passenger data is available in real-time, it is assumed reliable.

Moreover, it is possible that each line in the network is separately controlled by means of holding to achieve line-level regularity, but such potential holding decisions are ignored by the controller at the transfer stop. A consequence of this simplification is that if the vehicle of the connecting line, m_2, is held at intermediate stops, between its current location and the transfer stop, then the estimation of its arrival time at the latter might be underestimated. Therefore, if the controller decides that m_1 should be held to synchronise the transfer with m_2, then the internal controller of route r_2 should ensure that the estimated arrival time of m_2 at the transfer stop is not violated due to holding. Additionally, in order for this assumption to hold, the encounter of the connecting lines must be single-point, instead of flexible road-segment, as distinguished by Liu et al. [2014]. That is because, in case of the latter, the control decisions in the multiple transfer stops should be coordinated.

Another assumption is that the vehicles have unlimited capacity, hence any effects on passengers inflicted by the inability to board are ignored by the controller. Along with this comes the simplification that all passengers on-board experience the same level of comfort regardless of the vehicle occupancy and whether they are sitting or standing. Last but not least, the dwell time of a vehicle at a stop is considered to be independent of the passenger demand and consequently of the vehicle headway.

Passenger prediction model

As already mentioned in section 2.2, three data types regarding the passengers may be available in real-time, namely the vehicle occupancy, the passenger arrivals and the passenger destinations. These are to be referred to as 'APCs', 'tap-ins' and 'destination', respectively.

The prediction of the vehicle occupancy upon departure from a stop is a function of the on-board passengers when the vehicle arrives at that stop, the passenger demand there and

the number of alighting passengers. If the occupancy is available in real-time, then the occupancy upon arrival is known and the prediction concerns only the boarding and alighting volumes at that stop. In the absence of real-time data, the last known vehicle occupancy state is used, i.e. it was empty before starting its trip. That is sequentially updated for every stop visited by the vehicle.

$pocc_{m,s,r} =$		
$\hat{pocc}_{m,s-1,r} + pbd_{m,s,r} - pal_{m,s,r}$	if APCs are available	(1)
$\displaystyle\sum_{i=1}^{s} \left(pbd_{m,i,r} - pal_{m,i,r} \right)$	otherwise	

The number of boarding passengers equals the passenger demand at the stop, when the vehicle capacity is ignored and there is only one line at that stop.

$pbd_{m,s,r} =$	
$pdem_{m,s,r}$	(2)

The prediction of the passenger demand at a stop is a function of the passenger arrival rate and the vehicle headway from the previous vehicle of the same line. The headway can be defined as the difference of either the arrival or the departure times of the two vehicles. In this case, the arrival times are used, because they are independent of the control decision to be made, as opposed to the departure times, which contain also the dwell time at the stop and therefore depend on the holding time. This formulation, which according to Daganzo [2009] leads to simpler models, assumes that the vehicle collects only passengers arriving during the inter-arrival headway, ignoring those being generated while the vehicle is stopped.

When real-time data regarding passengers tapping in at a stop is available, this prediction can be improved. The detected number of passengers is used for the time interval between the departure of the previous vehicle and the current moment in time, while the prediction is applied only to the time interval between the present and the estimated arrival time of the following vehicle. If more lines make use of the same stop, then the passenger arrival rate may be given per line, yet the real-time detected passenger demand may have unspecified the line intended to be used. Therefore, an estimation method is needed in order to split the detected passengers among the existing lines at that stop. This split could, for instance, be based on the ratio of the arrival rate per line over the arrival rate of passengers at the stop.

$pdem_{m,s,r} =$		
$\hat{pdem}_{m,s,r} + \lambda_{s,r} * (tarr_{m,s,r} - tnow)$	if tap-ins are available	(3)
$\lambda_{s,r} * (tarr_{m,s,r} - tarr_{m-1,s,r})$	Otherwise	

When the destination of passengers tapping-in is known in real-time, then potential paths for this OD pair can be generated, followed by an assignment process. This assignment model should additionally be applied on passengers who are predicted to enter the network in the near future and whose destination is also an item for prediction. The latter is necessary for passengers already in the network in case that the destination is not available in real-time. Due to the complexity of this approach, as well as the scarcity of real-time destination data provision, an alternative method is adopted in the present study. The prediction of alighting and transferring passengers is, then, based on the alighting and transferring fractions per stop, respectively. These fractions depend on the vehicle occupancy, whose availability in real-time can improve the predictions.

$pal_{m,s,r} =$	
$pocc_{m,s-1,r} * \rho al_{s,r}$	(4)
$ptr_{m,s,i \rightarrow j} =$	
$pocc_{m,s,r} * \rho al_{s,r} * \rho tr_{s,i \rightarrow j}$	(5)

Depending on the types of data about the passengers that can be retrieved in real-time, the predictions regarding the passenger flows are made using the corresponding equations. These can then be used by the controller to determine the holding control strategy.

Holding control rules

The developed controller is activated once a vehicle enters a transfer stop and needs to decide what its dispatch time should be. The decision is made following a rule-based approach which considers a discrete set of candidate holding times. The controller compares the effect on the passengers inflicted by each of these times and selects and applies the one that is the least costly.

Prior to the discussion of the rules that govern the control process and the definition of the least costly alternative, two things need to be specified, namely the values to be considered in the choice set and the categories on which the effect on the passengers is measured.

The candidate holding times could correspond to the no-holding option and the one that permits the synchronisation of transfers. If multiple lines traverse the transfer stop in question, then a potential holding time is computed for every connecting line. Given the fact that the literature findings have demonstrated the benefits of applying holding to regularise the service, the holding time to achieve this regularity is also added to the choice set. The no-holding option may then be removed to reduce the computation time.

In order to synchronise the transfers, the controlled vehicle needs to be held until the arrival of the upcoming vehicle of the connecting line and provide sufficient time for the

transferring passengers to complete the transferring process. Applying this on the example of Figure 3.1, gives:

$$\text{thold}_{\text{syn}} =$$

$$\text{tarr}_{m_2,s_{12},r_2} - \text{tnow} + \tau tr_{s_{12},r_2 \to r_1} \tag{6}$$

The time needed for passengers to complete the transfer, $\tau tr_{s_{12},r_2 \to r_1}$, includes their alighting time, the time to walk to the connecting vehicle and also to board it. If the vehicle capacity were considered, then crowding effects would influence and increase the alighting and boarding times [van Oort et al., 2015]. However, under the assumption that capacity is ignored, the walking time between vehicles is the one defining the transfer time. The computation of the latter can make use of the classification of transfer types proposed by Hadas and Ceder [2010b]. According to them, there are three categories, namely the adjacent transfer points, where the connecting line is across the street, the nonadjacent transfer points, where passengers have to walk a certain distance to transfer, and the shared road-segment transfer points, where passengers just wait for the next vehicle. Depending on the configuration of the stop, the transfer type can be determined, along with the needed transfer time.

The holding time to restore service regularity is selected such that even headways between the preceding bus and the following bus are kept, while respecting the minimum headway requirement. This selection is justified by the superiority of this strategy in the literature [Cats et al., 2011]. If the hereby computed dispatch time has already elapsed, then the holding time is set to zero. For the example of Figure 3.1, this is:

$$\text{thold}_{\text{reg}} =$$

$$\max\left(\min\left(\left(\frac{\text{tarr}_{m_1+1,s_{12},r_1} + \text{tarr}_{m_1-1,s_{12},r_1}}{2}\right), \text{tarr}_{m_1-1,s_{12},r_1} + \alpha * \eta_{m_1-1,m_1}\right) - \text{tnow}, 0\right) \tag{7}$$

In both expressions for the calculation of the candidate holding times, the arrival time of vehicles at the transfer stop is needed. For the vehicles that have already traversed the stop (e.g. m_1-1), their arrival time is known and made available by the AVL data. In the case of approaching vehicles, however, their arrival time needs to be predicted. This prediction depends on their current location and the intermediate stops, and may be given by the summation of the expected running time in-between stops and the dwell time in each of them, assuming that these two are separately predicted. Alternatively, it may be based on the propagation of any existing delay to downstream stops, thereby estimating the arrival time by shifting the scheduled arrival time according to the expected delay.

The decision that is made by the controller affects three passenger streams, namely the passengers on-board, the passengers waiting at downstream stops and the passengers that intend to transfer from the connecting line. It is in this part of the process that the reliability

of the predictions regarding the passengers becomes valuable, since the more reliable the predictions are, the better the effect on each stream can be quantified.

The on-board passengers experience a delay while the vehicle is held:

$\textbf{Held}_y =$	
$\left(\text{pocc}_{m_1,s_{12}-1,r_1} + \text{pbd}_{m_1,s_{12},r_1} - \text{pal}_{m_1,s_{12},r_1}\right) * \text{thold}_y$	(8)

At the same time, the passenger arrival process at downstream stops of the controlled line still takes place and their waiting time is influenced as the vehicle is held upstream. The length of the horizon of downstream stops that are considered, μ, is of importance for this calculation. Also, in case that real-time data is available regarding the passenger tap-ins at stops, these real-time values should only be considered up to the stop that was last visited by the previous vehicle (for the given example that vehicle is m_1-1). The reason for that is that passengers downstream of that stop will be served by this previous vehicle and will not be affected by the control strategy applied on the currently controlled vehicle. At those stops, only the passengers that are going to be generated after the departure of that previous vehicle should be considered.

Due to the assumption of random passenger arrivals at the stops, their average waiting time equals half of the time interval during which they are generated prior to being served, while the passenger volume generated during this interval is the product of the passenger arrival rate at the stop and the length of the interval. When tap-ins are available, the arrival time of each detected passenger is known, along with the elapsed waiting time. Their remaining waiting time is the interval from now to the estimated arrival time of the controlled vehicle. During this interval more passengers are generated, to whom the average waiting time calculation applies.

$\textbf{DsWait}_y =$		
$\sum_{i=s_{12}+1}^{s_z} \left(\sum_{j \in \hat{p}\text{dem}_{m_1,i,r_1}} \left(\text{tnow} - \hat{p}\text{arr}_j\right) + \left(\text{tarr}_{m_1,i,r_{1y}} - \text{tnow}\right) * \hat{p}\text{dem}_{m_1,i,r_1} \right)$ $+ \sum_{i=s_{12}+1}^{s_z} \left(\frac{1}{2} * \left(\text{tarr}_{m_1,i,r_{1y}} - \text{tnow}\right)^2 * \lambda_{i,r_1} \right)$ $+ \left(\sum_{i=s_z+1}^{s_{12}+\mu} \frac{1}{2} * \left(\text{tarr}_{m_1,i,r_{1y}} - \text{tarr}_{m_1-1,i,r_1}\right)^2 * \lambda_{i,r_1} \right) * \delta$	if tap-ins are available	(9)
$\sum_{i=s_{12}+1}^{s_{12}+\mu} \frac{1}{2} * \left(\text{tarr}_{m_1,i,r_{1y}} - \text{tarr}_{m_1-1,i,r_1}\right)^2 * \lambda_{i,r_1}$	otherwise	

where:

s_z: last stop that vehicle $m_1 - 1$ departed from

$$\text{tarr}_{m_1,i,r_1\_y} = \begin{cases} \text{tnow} + \text{tdw}_{m_1,i,r_1} + \text{thold}_y + \text{tdr}_{i\to i+1,r_1} & i = s_{12} \\ \text{tarr}_{m_1,i-1,r_1\_y} + \text{tdw}_{m_1,i-1,r_1} + \text{tdr}_{i-1\to i,r_1} & i > s_{12} \end{cases}$$

$$\delta = \begin{cases} 1 & \text{if } s_{12} + \mu > s_z \\ 0 & \text{otherwise} \end{cases}$$

The transferring passengers can either board the vehicle directly upon their arrival, if the vehicle is held to provide transfer synchronisation, or they need to wait for the next vehicle of that line to arrive.

$\text{TrWait}_y =$		
$\text{ptr}_{m_2,s_{12},r_2\to r_1} * \left(\text{tarr}_{m_1+1,s_{12},r_1} - \text{tarr}_{m_2,s_{12},r_2}\right)$	if y = reg	(10)
0	if y = syn	

Each of these time components (Eq. (8)-(10)) reflects the effect of the control decision on a different group of passengers (on-board, downstream and transferring, respectively). They are brought together in a generalised cost function, based on which the comparison of the candidate holding times is executed. This cost function is composed of their weighted sum, since it has been shown by value of time studies, that the passengers perceive differently each of these time components. For example, Wardman [2004] found that the ratio between the waiting time at stops and the in-vehicle time is in the range of 1.5-2.

$\text{Cost}_y =$	
$\beta_1 * (\text{TrWait}_y + \text{DsWait}_y) + \beta_2 * \text{Held}_y$	(11)

In case that the next vehicle (m_1+1) is expected to arrive after the vehicle of the connecting line (m_2), the transfer synchronisation should be considered for m_1+1, when it reaches the stop and activates the controller, and m_1 should only be held to restore regularity. Otherwise, the controller has to decide if the additional holding time for the transfer synchronisation should be allocated. This decision is based on the comparison of the generalised costs of the alternatives, selecting the one with the lowest cost.

More formally, the aforementioned rules are described by the following pseudocode:

IF $\text{tarr}_{m_1+1,s_{12},r_1} > \text{tnow} + \text{thold}_{\text{syn}}$ THEN
    IF $Cost_{\text{syn}} \leq \text{Cost}_{\text{reg}}$ THEN
        $\text{thold} = \text{thold}_{\text{syn}}$
    ELSE
        $t\text{hold} = \text{thold}_{\text{reg}}$
    END IF
ELSE
    $\text{thold} = \text{thold}_{\text{reg}}$
END IF

3.1.2 C^2_PassIT: PassIT with capacity constraints

This section describes the introduction of vehicle capacity constraints in the controller and this new variant is to be referred to as C^2_PassIT. The inclusion of vehicle capacity in the controller improves the estimate of the number of passengers that can board a vehicle under highly saturated conditions, by restricting it to the minimum of the passenger demand at the stop and the remaining space in the vehicle. The latter is the capacity reduced by the occupancy after passengers have alighted at the stop.

$$\mathbf{pbd_{m,s,r}} = \min\left(pdem_{m,s,r}, \kappa_m - pocc_{m,s-1,r} + pal_{m,s,r}\right) \tag{12}$$

Passengers that are unable to board due to lack of space in the vehicle are left behind at the stop and need to wait for the next vehicle to arrive. One assumption that is made for those left-behind passengers is that all of them will remain at the stop and wait for a vehicle of the line that they can board.

$$\mathbf{plb_{m,s,r}} = pdem_{m,s,r} - pbd_{m,s,r} \tag{13}$$

At the transfer stop more passengers may be generated for the controlled vehicle, while it is being held. Even though they can be assumed to have a negligible on-board delay, these passengers further reduce the available space in the vehicle and need to be considered. The transferring passengers that cannot board the vehicle due to capacity constraints need to wait for the next vehicle (m_1+1), so even in the case of transfer synchronisation it is possible that some passengers experience transfer waiting time due to denied boarding. The remaining space of m_1+1 is also important to be estimated, because, if the transferring passengers cannot fit, their transfer waiting time will increase.

$$\mathbf{ptrf_y} = \begin{cases} \max\left(ptr_{m_2,s_{12},r_2 \to r_1} - \kappa_{m_1+1} + pocc_{m_1+1,s_{12},r_1}, 0\right) & \text{if } y = reg \\ \max\left(ptr_{m_2,s_{12},r_2 \to r_1} - \kappa_{m_1} + pocc_{m_1,s_{12},r_1} + \lambda_{s_{12},r_1} * thold_y, 0\right) & \text{if } y = syn \end{cases} \tag{14}$$

In order to reduce the complexity of the controller, two assumptions are made with respect to the vehicle m_1+2. The first is that any transferring passengers left behind by vehicles m_1 and m_1+1 will be able to board m_1+2, so its remaining space is not computed. The second is that the arrival time of m_1+2 at the transfer stop will adhere to the desired headway of the line. These assumptions are reasonable since their violation would indicate either a severe network disruption or a poor design of the network supply, both of which go beyond the scope of the devised controller. Based on these assumptions, the transfer waiting time due to denied boarding can be estimated.

$\text{TrDenied}_y =$		
$\text{ptrf}_y * \eta_{m_1+1,m_1+2}$	if y = reg	(15)
$\text{ptrf}_y * \left(\text{tarr}_{m_1+1,s_{12},r_1} - \text{tarr}_{m_2,s_{12},r_2}\right)$ $+\max\left(\text{ptrf}_y - \kappa_{m_1+1} + \text{pocc}_{m_1+1,s_{12},r_1}, 0\right) * \eta_{m_1+1,m_1+2}$	if y = syn	

The estimation of the waiting time of passengers downstream the transfer stop is also affected by the introduction of capacity constraints. In addition to the randomly generated passengers at those stops, passengers may be left behind by vehicle m_1-1, thereby increasing the demand for m_1. In the absence of real-time passenger data, the number of passengers left behind can be estimated by tracing the occupancy and the expected demand for m_1-1 within the considered horizon. Their elapsed waiting time is taken as half of the time interval between the arrivals of m_1-2 and m_1-1.

When tap-ins are available, the number of passengers left behind by m_1-1 at stops downstream of the transfer stop is known, along with the arrival time of each passenger. At stops further downstream which m_1-1 has yet to visit, the number of passengers left behind and their waiting time need to be estimated. Having the number of tap-ins in real-time can lead to a better estimate of the demand, and for those passengers already generated their arrival time is known. The arrival time for those that still need to be generated is assumed to be at half of the time interval between now and the expected arrival of m_1-1. The passenger boarding process is assumed to follow a first-in-first-out (FIFO) regime, i.e. passengers who arrive first are the first to board. The passengers who are left behind are, hence, those who last arrived at the stop. Their waiting time is, then, the time they already have spent at the stop plus the time till m_1 arrives there.

Another effect created downstream relates to the controlled vehicle (m_1) itself. Its occupancy is increased during its holding time by passengers arriving within that time and by the transferring passengers, when the transfer is synchronised. This reduces the available space of m_1 and may result in leaving passengers behind at downstream stops, who need to wait for m_1+1.

The waiting time of downstream passengers (Eq. (9)) is split into two components, reflecting the time they waited for the first vehicle to arrive and the additional waiting time for those who were denied boarding by the first vehicle until their boarding time. The reason for this split is that the perception of this additional time by passengers is higher [Cats et al., 2016].

$\mathbf{DsWait_y} =$		
$$\begin{aligned} &\sum_{i=s_{12}+1}^{s_{12}+\mu}\left(\sum_{j\in(plb_{m_1-1,i,r_1}\cap\hat{p}dem_{m_1,i,r_1})}\left(tarr_{m_1-1,i,r_1}-\hat{p}arr_j\right)\right)\\ &+\sum_{i=s_{12}+1}^{s_z}\left(\sum_{j\in(\hat{p}dem_{m_1,i,r_1}-plb_{m_1-1,i,r_1})}\left(tarr_{m_1,i,r_{1y}}-\hat{p}arr_j\right)\right)\\ &+\sum_{i=s_{12}+1}^{s_z}\left(\frac{\left(tarr_{m_1,i,r_{1y}}-tnow\right)}{2}*\left(pdem_{m_1,i,r_1}-plb_{m_1-1,i,r_1}\right)\right)\\ &+\delta*\sum_{i=s_z+1}^{s_{12}+\mu}\frac{\left(tarr_{m_1,i,r_{1y}}-tarr_{m_1-1,i,r_1}\right)}{2}*\left(pdem_{m_1,i,r_1}-plb_{m_1-1,i,r_1}\right) \end{aligned}$$	if tap-ins are available	(16)
$$\begin{aligned} &\sum_{i=s_{12}+1}^{s_{12}+\mu}\left(\frac{\left(tarr_{m_1,i,r_{1y}}-tarr_{m_1-1,i,r_1}\right)}{2}*\left(pdem_{m_1,i,r_1}-plb_{m_1-1,i,r_1}\right)\right)\\ &+\sum_{i=s_{12}+1}^{s_{12}+\mu}\left(\frac{\left(tarr_{m_1-1,i,r_1}-tarr_{m_1-2,i,r_1}\right)}{2}*plb_{m_1-1,i,r_1}\right) \end{aligned}$$	otherwise	
$\mathbf{DsDenied_y} =$		
$$\begin{aligned} &\sum_{i=s_{12}+1}^{s_{12}+\mu}\left(\left(tarr_{m_1,i,r_{1y}}-tarr_{m_1-1,i,r_1}\right)*plb_{m_1-1,i,r_1}\right)\\ &+\sum_{i=s_{12}+1}^{s_{12}+\mu}\left(\left(tarr_{m_1+1,i,r_1}-tarr_{m_1,i,r_{1y}}\right)*plb_{m_1,i,r_1}\right) \end{aligned}$$		(17)

The generalised cost function (Eq. (11)) is then replaced by:

$\mathbf{Cost_y} =$	
$\beta_1*(TrWait_y+DsWait_y)+\beta_2*Held_y+\beta_3*(TrDenied_y+DsDenied_y)$	(18)

Upstream of the transfer stop the passengers that have been left behind by previous vehicles are independent of the control decision, yet their consideration may lead to better estimates of the remaining space in the vehicles of the controlled route, as well as a better estimate of the occupancy of m_2 and, consequently, the number of transferring passengers. In the absence of real-time tap-ins, which would already contain the left-behind passengers, their volume needs to be estimated by tracing the occupancy of the vehicles m_1-1 and m_2-1. Since those vehicles have already moved beyond the transfer stop, they have been controlled by the controller, which enables the creation of an internal memory with respect to the calculations made during previous control decisions. The occupancy of the previously

controlled vehicle is also useful to store in memory, because it is needed for the calculation of the number of passengers left behind at the downstream stops. Retrieving this memory saves computational time and reduces the complexity of the controller.

The last modification in the controller concerns the case when the vehicle occupancy is the only available real-time data source about the passengers. In that case, the calculation of potential left-behind passengers has to take place only at those stops where the vehicle departed full. At those stops, the available space upon arrival is known through the occupancy at the previous stop and the passenger demand at the stop can be estimated. The difference between the expected demand and the available space gives the number of passengers left behind there.

3.1.3 C³_PassIT: C²_PassIT with on-board crowding

In this third variant of the controller (C³_PassIT), the vehicle capacity is divided into seated and standing passengers so that the difference in their comfort level due to on-board crowding is taken into consideration in the cost function.

The controller assumes that while there are seats available, passengers will occupy them, thereby having the highest possible level of comfort. Once the seat capacity of the vehicle is depleted, then the remaining passengers have to stand. Moreover, those on-board have the priority over boarding passengers in acquiring a seat.

$\text{pseat}_{m,s,r} = $ $$\min\left(\text{pocc}_{m,s,r}, \varphi_m\right)$$	(19)
$\text{pstand}_{m,s,r} = $ $$\max\left(\text{pocc}_{m,s,r} - \varphi_m, 0\right)$$	(20)

While standing, passengers experience a lower comfort level which translates into a higher perception of the time they spend on-board. This effect is taken into account during the time that the vehicle is held at the stop, as well as while it is driving downstream of the transfer stop within the considered horizon.

The held time of on-board passengers (Eq. (8)) is then computed by:

$\text{Held}_y = $ $$\left(\gamma_1 * \text{pseat}_{m_1,s_{12},r_1} + \gamma_2 * \text{pstand}_{m_1,s_{12},r_1}\right) * \text{thold}_y$$	(21)

The time that passengers perceive to spend inside a vehicle is the summation of the time spent in-between stops, weighted by the amount of passengers standing or sitting and their respective on-board crowding multiplier.

$$IvT_{m_y} =$$

$$\sum_{i=s_{12}+1}^{s_{12}+\mu} \left(\gamma_1 * pseat_{m,i,r} + \gamma_2 * pstand_{m,i,r}\right) * \left(tarr_{m,i,r_y} - tarr_{m,i-1,r_y}\right) \quad (22)$$

This calculation is performed for the controlled vehicle (m_1) but also for its successor (m_1+1). That is because the former alone will always penalise the longer waiting time which serves more passengers and increases the vehicle occupancy. By computing the perceived in-vehicle time for m_1+1 and taking the difference of the two values, the effect of having uneven demand distribution along the vehicles of a line is estimated and may affect the selection of the control strategy. The absolute difference is considered, since evening out the demand distribution requires the difference to approach a value close to zero and not a value as small as possible. These modifications result in the following generalised cost function:

$$Cost_y =$$

$$\beta_1 * (TrWait_y + DsWait_y) + \beta_2 * Held_y + \beta_3 * (TrDenied_y + DsDenied_y)$$
$$+ \beta_4 * \left| IvT_{m_{1_y}} - IvT_{m_1+1_y} \right| \quad (23)$$

3.1.4 UC³_PassIT: C³_PassIT with demand level uncertainty

The last controller variant introduces uncertainty in the passenger flow estimates. In the previous controllers, the passenger generation and alighting processes have been assumed to be deterministic, based on average arrival and alighting rates, respectively. However, these processes are stochastic and a new variant of the controller is developed to take stochasticity into account. This variant is referred to as UC³_PassIT.

The number of arriving passengers is assumed to follow the Poisson distribution, which is an approach adopted by many models in the literature for short-headway services [Cats, 2011; Delgado et al., 2012; Dessouky et al., 2003; Fu and Yang, 2002; Hans et al., 2015]. The parameter of the distribution is equal to the product of the average passenger arrival rate and the time during which the passenger generation is estimated. The latter is the vehicle headway when no real-time data is available, while in case that tap-ins are available, it is the difference between the estimated vehicle arrival and the current time.

$pdem_{m,s,r} \sim$		
$Poisson\left(\lambda_{s,r} * (tarr_{m,s,r} - tnow)\right)$	if tap-ins are available	(24)
$Poisson\left(\lambda_{s,r} * (tarr_{m,s,r} - tarr_{m-1,s,r})\right)$	Otherwise	

The number of alighting passengers is assumed to vary according to the Binomial distribution, with parameters the vehicle occupancy upon arrival at the stop s and the average alighting rate at that stop.

$$\mathbf{pal_{m,s,r}} \sim$$

$$\text{Binomial}\big(\text{pocc}_{m,s-1,r} \, , \rho al_{s,r}\big) \tag{25}$$

The Binomial distribution is also applied on the estimation of the number of transferring passengers, where the parameters are the estimated number of alighting passengers at the stop and the average transferring rate from one line to another.

$$\mathbf{p\_tr_{m,s,i \to j}} \sim$$

$$\text{Binomial}\big(p\_al_{m,s,r} \, , \rho\_tr_{s,i \to j}\big) \tag{26}$$

Having defined these distributions, the estimate of each of these variables requires the specification of the percentile of the distribution on which the demand level is computed. Given a percentile, the predictions of the passenger flows can be executed and a holding control strategy is decided based on the holding control rules described in 3.1.1. Since it is uncertain at which level the demand lies, a sample of several percentiles is drawn and a control decision is taken for each of them.

Ideally, there would be a cut-off point, beyond which the decision switches from holding to regularise towards holding to synchronise, or vice versa. However, due to the interdependencies of the variables and their discrete nature, this is not the case. Since the control decision is binary, there are two ways to combine the individual decisions made for each percentile. The first would be to use the frequency with which each decision is made throughout the sample. Yet, this would require a big sample to ensure that the probability of occurrence of each draw, and, consequently, the weight of each decision, is equal. The second way would be to add up the costs of each holding option for every draw and when the sample is exhausted, make a decision based on the comparison of the two total costs. Using this approach, the binary character of the decision would be applied only in the final decision. Because of this property of the second method, it is the one selected in the present study.

3.1.5 Controller comparison

In this section, four real-time transfer synchronisation controllers have been developed which make use of different passenger data sources. All of them apply a rule-based strategy that chooses the holding time which yields the lowest total generalised passenger cost. Four passenger data sources are distinguished: (1) no real-time passenger data; (2) real-time vehicle occupancy; (3) real-time passenger tap-ins; (4) combination of (2) and (3).

The four controllers differ in complexity and in their underlying assumptions. The first controller, PassIT, assumes unlimited vehicle capacity. C^2_PassIT adds the consideration of capacity constraints for the transit vehicles, thereby restricting the number of passengers that can board, potentially leaving some of them behind for the next vehicle of the line. C^3_PassIT further increases the complexity by taking into account the effect of the on-board crowding

on the passenger level of comfort. The last variant, UC³_PassIT, introduces uncertainty in the passenger demand levels by drawing values from a distribution instead of using the average values. A comparison of the components included by each controller in the passenger flow predictions and the generalised cost calculations is shown in Table 3.1.

Table 3.1: Controller comparison.

Controller		PassIT	C² PassIT	C³ PassIT	UC³ PassIT
Passenger flow predictions	Demand generation	x	x	x	x
	Alighting	x	x	x	x
	Transferring	x	x	x	x
	Denied boarding		x	x	x
	Seated			x	x
	Standing			x	x
	Prediction uncertainty				x
Passenger cost components	Waiting (Transferring and Downstream)	x	x	x	x
	Held (On-board at transfer stop)	x	x	x	x
	Denied boarding (Transferring and Downstream)		x	x	x
	Riding (On-board crowding conditions)			x	x

3.2 Simulation model

The simulation model used in this study is BusMezzo. It enables the analysis and evaluation of holding control strategies which are the control methods of interest for the present study. Yet, the control rules implemented within BusMezzo differ from those of the controllers described in section 3.1. The latter have been implemented in MATLAB and, therefore, need to be coupled to the simulation model.

This section first describes the features of BusMezzo that are relevant to this study (subsection 3.2.1) and then explains the coupling process between the simulation model and the controllers (subsection 3.2.2).

3.2.1 BusMezzo

BusMezzo is a mesoscopic, dynamic transit operations and assignment model [Cats et al., 2016; Toledo et al., 2010]. It is an event-based simulator, which simulates the progress of transit vehicles and passengers using an agent-based approach. It models individual vehicles, but does not represent lanes explicitly. This mesoscopic level is desired, because it avoids the detailed modelling of the second-by-second vehicle movements which would have limited the application scope to the level of a road section. Previous studies have demonstrated that the model can reproduce the bunching phenomenon [Cats et al., 2010] and represent dynamic congestion effects including variations in on-board crowding and denied boarding [Cats et al., 2016].

The scheduled travel time between stops is specified, so that the lateness or earliness can be computed upon arrival at each stop. This is used in the prediction of the arrival time at downstream stops, in order to capture the delay propagation through the network. The interaction with motorised traffic and congestion effects are modelled in the actual travel time between stops, which can be given as a distribution from which values are stochastically drawn. At the stops, the dwell time is a function of the passenger activity at the stop, the on-board crowding and the physical characteristics of the stop, such as the amount of vehicles it can fit and whether it is a bay stop or not.

Along the transit routes, any number of stops may be defined as control points, where a holding strategy determines the dispatching time of the vehicle. Passengers arriving during the holding time can board the vehicle, while capacity constraints on transit vehicles are explicitly modelled, so that passengers who are denied boarding due to overcrowded conditions have to wait for the next vehicle.

The assignment model is agent-based, enabling the behavioural modelling of individual travellers. Each passenger is modelled as an adaptive decision maker, whose choices are based on random utility discrete choice models. The demand is given as an OD matrix and each traveller can choose among the available paths in the choice set for each OD pair, which may be dynamically updated throughout the trip as more information becomes available, shaping the expectations.

It is worth noting that the simulation allows for a warm-up and a cool-down period for the circulation of the vehicles and passenger demand is only generated in-between those. At the beginning of the simulation, the warm-up period ensures that transit vehicles will be distributed over the transit network when passengers start arriving, thereby computing reasonable waiting times for the passengers. Towards the end of the simulation, the cool-down period serves the purpose of letting all passengers reach their destination, thus having a finite total travel time for their trip.

The holding control strategies within BusMezzo allow for scheduled-based holding, as well as headway-based. In the latter, the headway from the preceding vehicle may be considered, or from the next one, or their combination, as the mean of the two headways. The last strategy that is available sets an upper limit to this mean headway by allowing a holding time up to the planned headway. This is similar to the calculation of the holding time for regularity in the controllers described in section 3.1. Since these strategies ignore the possibility to synchronise transfers between lines and do not take into account the location of passengers in the network, the developed controllers are coupled to BusMezzo to determine the holding time.

3.2.2 Coupling process

The simulation model is used for mimicking real-world operations as a testbed for testing the performance of the controller and its consequences under different scenarios. Each time a transit vehicle enters a transfer stop, BusMezzo calls the control instance in MATLAB. Figure 3.2 presents schematically the data exchange between the simulation model and the PassIT controller. Figure 3.3 does the same for the rest of the controller, which are grouped under the name of X_PassIT. The simulation model provides the controller with real-time vehicle and passenger data as well as predictions regarding vehicle arrival times at downstream stops within a predetermined horizon. These outputs are fed into the controller where the passenger data exchange depends on the assumed data availability.

The controller inputs and outputs shown in these figures are described in the ensuing, along with the modifications made to the controller in order to facilitate this coupling.

Controller inputs

When a vehicle arrives at a transfer stop, s_{12}, which is defined to be a control point of that line, the controller is activated. Since the objective of the proposed controller pertains to the passenger travelling experience, it is only used when passengers are present in the network. Therefore, it remains inactive during the warm-up and cool-down periods of the simulation. In the rest of the simulation time, upon the arrival of a vehicle and prior to any passenger activity (i.e. boarding and alighting) at the transfer stop, this local instance is created and sent to the controller. It includes data about the vehicles, as well as the passengers in the network.

The vehicle activity is represented by their arrival time, which is either recorded or predicted. The former corresponds to stops that the vehicles have already visited, from the beginning of the line (stop #1) till their present location. For stops further downstream which are within the designated horizon, a prediction of their arrival time at each of them is made and sent to the controller.

The vehicles considered by each controller along with their corresponding horizon are shown in Table 3.2. It should be noted that the UC^3_PassIT is not displayed in the table since it is identical to C^3_PassIT with respect to the depicted properties. Following the notations of the example of Figure 3.1, PassIT requires the arrival times of the vehicles m_1 and m_1-1 from the beginning of the line and up to μ stops downstream the transfer stop, and for vehicles m_2 and m_2-1 from the beginning of the line and up to the transfer stop. The arrival time of m_1+1 is only of interest at the transfer stop. The X_PassIT controllers additionally need the arrival times of m_1+1 from the beginning of the line and up to μ stops downstream the transfer stop and of m_1-2 from the transfer stop up to μ stops downstream.
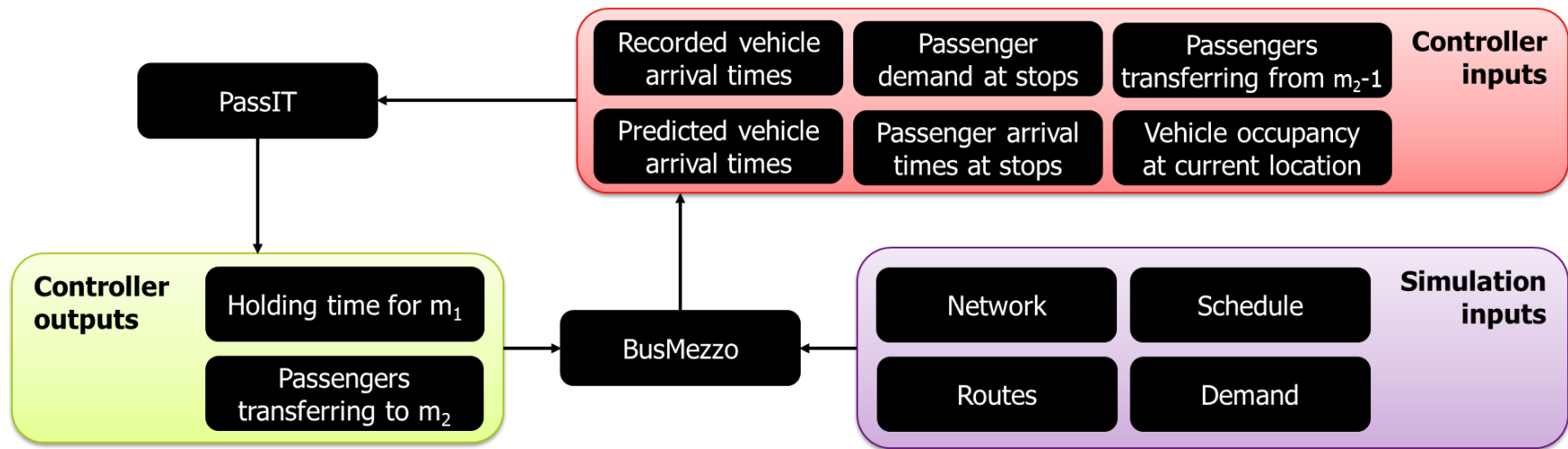
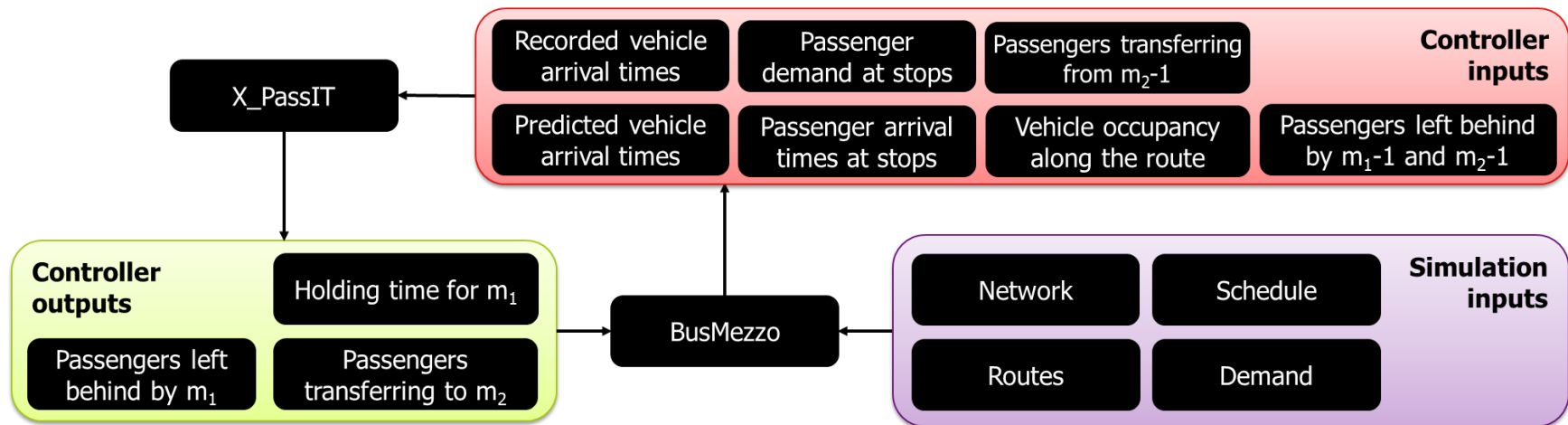Figure 3.2: Schematic relations between the PassIT controller and the BusMezzo simulation model.

Figure 3.3: Schematic relations between the other controllers (X_PassIT) and the BusMezzo simulation model.

Table 3.2: Stops considered by each controller regarding the vehicle arrival time and occupancy.

Vehicle	Arrival time horizon			Occupancy horizon		
	PassIT	c^2_PassIT	c^3_PassIT	PassIT	c^2_PassIT	c^3_PassIT
m_1	1 up to $s_{12}+\mu$	1 up to $s_{12}+\mu$	1 up to $s_{12}+\mu$	1 up to $s_{12}+\mu$	1 up to $s_{12}+\mu$	1 up to $s_{12}+\mu$
m_1+1	at s_{12}	1 up to $s_{12}+\mu$	1 up to $s_{12}+\mu$		1 up to s_{12}	1 up to $s_{12}+\mu$
m_1-1	1 up to $s_{12}+\mu$	1 up to $s_{12}+\mu$	1 up to $s_{12}+\mu$		s_{12} up to $s_{12}+\mu$	s_{12} up to $s_{12}+\mu$
m_1-2		s_{12} up to $s_{12}+\mu$	s_{12} up to $s_{12}+\mu$			
m_2	1 up to s_{12}	1 up to s_{12}	1 up to s_{12}	1 up to s_{12}	1 up to s_{12}	1 up to s_{12}
m_2-1	1 up to s_{12}	1 up to s_{12}	1 up to s_{12}			

With respect to the passengers, BusMezzo sends the vehicle occupancy, the amount of tap-ins at the stops and the arrival time of each of the waiting passengers there. Depending on the level of real-time passenger data that is assumed available by the controller, a subset of these is used. Four levels are defined: (1) no real-time passenger data; (2) real-time vehicle occupancy; (3) real-time passenger tap-ins; (4) combination of (2) and (3).

In the absence of real-time data, the vehicle occupancy that is of interest per vehicle and controller is shown in Table 3.2. In level (2), PassIT requires the occupancy of vehicles m_1 and m_2 at their current location, because those transport the passenger flows of interest for the cost calculations (held on-board and transfer waiting time). X_PassIT needs to track the load of m_1-1 and m_1+1 as well, due to the consideration of capacity constraints. Improved estimates of the occupancy of the former can result in better estimates of how many passengers it may leave behind at the downstream stops. Regarding m_1+1, its estimated occupancy upon arrival at the transfer stop can determine how many transferring passengers of the m_2 can fit in case that the transfer is not synchronised or if the remaining space of m_1 cannot fit them all. Moreover, X_PassIT makes use of the recorded vehicle occupancy from the beginning of the line up to their current location. The reason for this discrepancy is the need to estimate the amount of passengers that may have been left behind due to capacity constraints.

In level (3), the real-time passenger demand reflects the demand that has been generated at downstream stops within the horizon of each vehicle, as well as the demand that had been recorded when each of these vehicles visited the stops upstream of their current location. The recorded tap-ins compensate for the lack of real-time data regarding the vehicle occupancy, since by knowing how many passengers wanted to board at every stop, the estimate of the vehicle occupancy can be improved.

In Figure 3.4 an example of the expected vehicle occupancy along the route for each of the four real-time passenger data levels is displayed. As already explained by Eq. (1), the estimation of the occupancy is a function of the amount of passengers boarding and alighting and of the occupancy at the previous stop. When it is known in real-time, the estimation of the boarding and alighting passengers is not necessary, except for in the case that it equals its capacity upon departure from a stop. In this case, it is possible that passengers have been left behind at the stop and their volume should be estimated. However, it is known up to the last visited stop and it needs to be predicted for stops downstream. This prediction differs depending on whether tap-ins are available or not. When tap-ins are available (level (3) and (4)), the amount of passengers expected at downstream stops is only partially known, since by the time the vehicle arrives there more passengers may have been generated. The white bars in the figure represent the amount of alighting passengers, which are subtracted from the last estimated occupancy, leaving the coloured bar to represent the vehicle load.
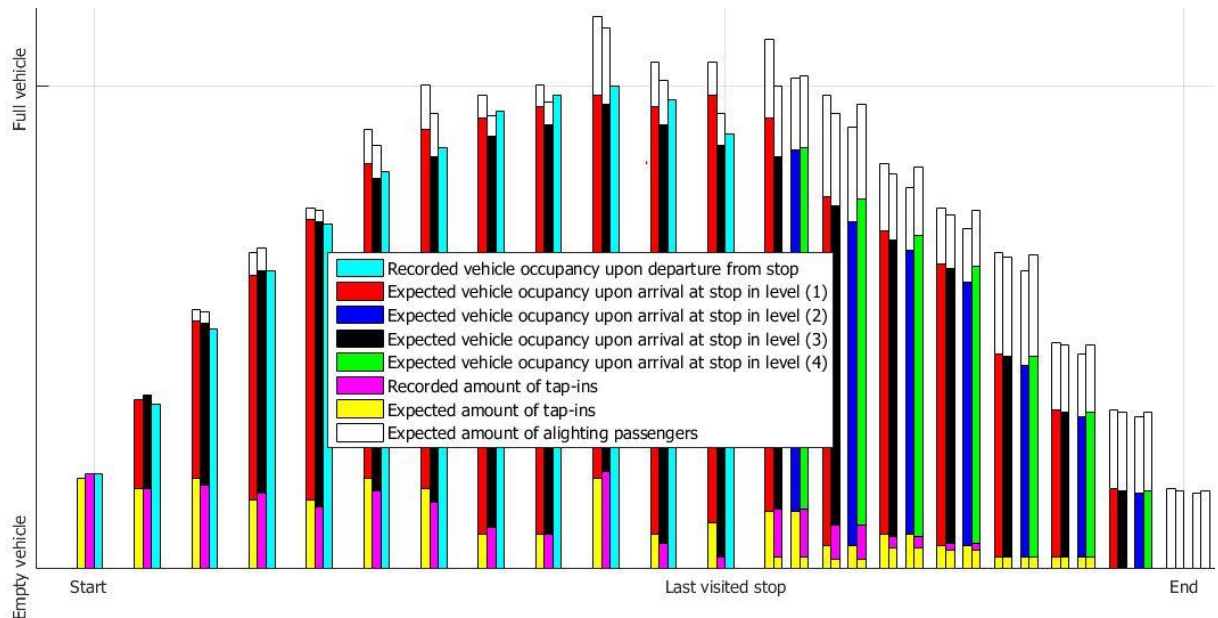
Figure 3.4: Expected vehicle load along the route for each of the four passenger data levels.

At downstream stops, the amount of passengers already generated is given as input to the controller in level (3), along with the arrival time of each passenger. An example of the passenger demand estimation at a downstream stop is shown in Figure 3.5. When no real-time data is available regarding the tap-ins, the expected amount of passengers is linearly estimated between successive vehicle arrivals based on the historic average of the passenger arrival rate at that stop. When tap-ins are known, the amount and arrival time of each passenger is given up to the current time (marked by asterisks in the figure) and the demand estimation applies for the remaining time until the arrival of the next vehicle. This serves the purpose of a better estimation of the downstream waiting time in the cost function, which is given by the area under the corresponding curve.
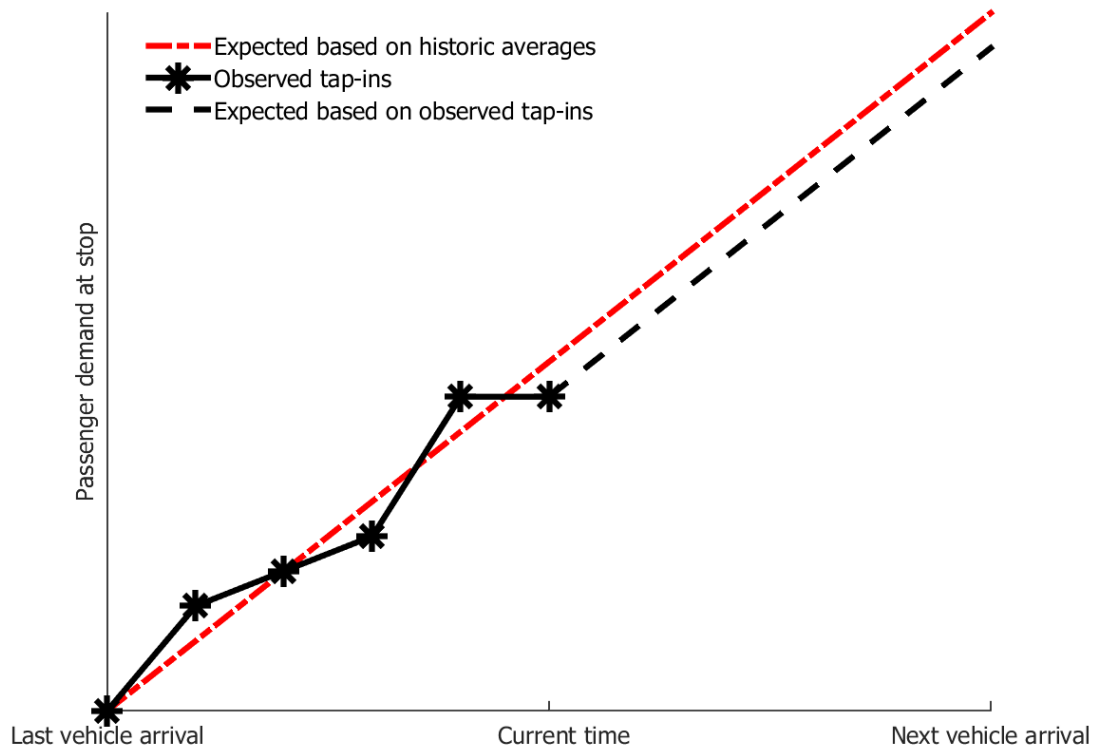
Figure 3.5: Expected passenger demand at a stop between successive vehicle arrivals.

The demand that is detected at the transfer stop may be headed for either of the lines. Therefore, the controller needs to estimate the share of the real-time demand that intends to board m_1. As suggested in subsection 3.1.1, this estimate is based on the ratio of the passenger arrival rate for m_1 over the total arrival rate of passengers at the transfer stop. Another peculiarity with respect to the demand detected at the transfer stop is that it may contain passengers transferring from one line to the other. Their volume should be estimated and subtracted from the detected volume prior to the calculation of the share per line.

Controller outputs

The outputs of the controller serve two different functions, namely the continuation of the simulation run and the acceleration of future control decisions. The latter is achieved by storing some variables in memory and retrieving them later as input, thereby avoiding to repeat the same calculations that led to them, which would necessitate even more inputs.

As previously mentioned, in the controller inputs description, the amount of passengers who have alighted at the transfer stop from a vehicle of the other line that previously visited the stop (i.e. m_2-1) should be estimated and distinguished from the demand that was generated there. This volume estimation is based on the vehicle occupancy upon arrival of m_2-1 at the transfer stop, the alighting rate and the transferring rate. It is computed when m_2-1 is being controlled and saved as an output. The implementation of the memory function involves the communication of this saved value, in the form of an input to the vehicle that visits the stop next.

Obviously, if m_2-1 had taken as a control decision to synchronise the transfer with m_1, then the passenger flow from m_1 to the other line would be able to board m_2-1 immediately, at least partially, depending on the remaining space in the vehicle. This affects the estimate of the number of passengers that will have to wait for the m_2 vehicle. In the event that m_1+1 arrives before m_2, the estimated volume of transferring passengers from m_1+1 is added to those already waiting to transfer from m_1.

A similar memory function is performed for three more passenger-related variables of the X_PassIT controllers. In levels (1) and (2), where the tap-ins are not available, the number of passengers that were estimated to have been left behind by m_1-1 and m_2-1 can improve the estimation of the passenger demand at the stops. Therefore, they are saved when the respective vehicle is at the transfer stop and sent back to the controller when m_1 is being controlled. The third variable concerns the occupancy of m_1-1 when it departed from the transfer stop and it is useful for levels (1) and (3) which do not have real-time occupancy data available. By having this estimate, the progress of its occupancy estimation downstream the transfer stop can start at the first downstream stop, avoiding the repetition of the calculations from the beginning of the line.

Apart from these memory-creation outputs, the controller decides whether the vehicle should be held for synchronisation or to restore regularity. The amount of the holding time that corresponds to each alternative is estimated based on the vehicle arrival time predictions and depending on the decision that is made, the dispatch time of the vehicle is determined. In the case of a decision to synchronise the transfer, the holding time that was estimated, does not bind the vehicle dispatch time. Instead, the vehicle is instructed to wait till another vehicle arrives at the stop. If m_2 arrives first then m1 dispatches once the transferring passengers have boarded it, while in case m_1+1 arrives before m_2, m_1 dispatches immediately and a new control decision is made for m_1+1.

The motivation behind this implementation is that the arrival time of upcoming vehicles is merely a prediction and can therefore be wrong. An underestimation of the arrival time of m_2 would lead to the holding of m_1 for less time that actually needed to synchronise the transfer, while an overestimation would keep the bus longer than necessary at the transfer stop. Even though direct transfers are ensured by forcing the vehicle to wait till the other one actually arrives, there is a downside to this approach. The decision to synchronise was made for a holding time less than the realised one, which means that it is possible that the controller would have made a different decision with hindsight.

Controller modifications

Due to the fact that the vehicle arrival time predictions are not fully reliable, as assumed during the development of the controllers, a number of modifications are necessary to prevent irrational outcomes.

The first one relates to the passenger demand estimation, which is a function of a time interval (either between successive vehicle arrivals or the present time and the estimated vehicle arrival time). Given that a wrong prediction could state that a vehicle should have already arrived at a stop, i.e. prior to the present time, the interval can get negative values, which would result in negative passengers being generated. In order to prevent this, any negative time interval is zeroed, thereby zeroing the expected passenger arrivals till the vehicle actually arrives at the stop.

Another modification is applied to the estimation of the holding time for regularity. Its original calculation makes use of the vehicle arrival headway. However, when m_1 is held for synchronisation and m_1+1 arrives before m_2, their arrival headway could designate that the service is regular, which is unrealistic, since the two successive vehicles are bunched at the transfer stop. In this case, the holding time for regularity of m_1+1 is recalculated, using the present time in place of the arrival time of m_1. Moreover, the arrival time prediction of $m1+1$ at downstream stops is re-evaluated for the regularity control case based on the new holding time for regularity.

Last but not least, and once again for the case that m_1+1 arrives before m_2 when m_1 had decided to synchronise, the arrival time prediction of m_1 at downstream stops needs to be adjusted. The original predictions were based on the arrival time of m_1 at the transfer stop and a different holding time than the one actually implemented, so they need to be offset by the difference between the expected and the realised holding time.

3.3 Performance evaluation

At the end of a simulation run, BusMezzo generates output files that summarise the progress of vehicles and the passenger activity in the network throughout the simulation. They are given both on an individual basis, as well as aggregated per stop or line. By processing these outputs, the performance can be evaluated. This evaluation is based on three pillars, namely the performance measures (subsection 3.3.1), the reference case (subsection 3.3.2) and the number of runs (subsection 3.3.3). Each of them is further explain in this section.

3.3.1 Performance measures

The performance measures that are used to evaluate holding control strategies have been reviewed in the subsection 2.1.4. Since the objective of the proposed controllers pertains to the passenger travelling experience (i.e. minimising their perceived travel time), and especially looks into the synchronisation of transfers, it is reasonable to select key performance indicators (KPIs) concerning the passenger activity.

A primary objective is the improvement of the passenger travelling experience. As an indicator of that, the first KPI to be used, is the perceived trip time from origin to destination per passenger. The time components considered, include the waiting time at the origin, the driving time in the vehicle and if applicable, the time waiting time at the transfer stop and the

time spent on-board a vehicle that was held at a control stop. The waiting time components are further split to distinguish between the time that the passenger waited for the first vehicle arrival at the stop and the time till that passenger boarded a vehicle. If the passenger boarding time is greater than the departure time of the first vehicle, then the passenger was not able to board the first vehicle due to capacity constraints and needs to wait for the next one. The reason for this discrimination is that the perception of the time after being denied boarding is significantly higher than the time waiting for the first vehicle to arrive [Cats et al., 2016]. Additionally, this perceived time is evaluated separately for four passenger groups, since the effect of the control strategy differs among them. These groups are the following:

a) Upstream: passengers whose destination is upstream the transfer stop.
b) Downstream: passengers whose origin is downstream the transfer stop.
c) Transferring: passengers who change lines at the transfer stop.
d) Traversing: passengers whose origin is upstream the transfer stop and whose destination is downstream the transfer stop.

Since the devised controllers envisage the reduction in the transfer waiting time, the distribution of the latter can indicate to which extent this goal is achieved. The transfer waiting time is defined as the difference between the arrival times at the transfer stop of the two vehicles passengers used in their route. If their second vehicle arrived first, then their waiting time is considered to be zero.

According to the analysis in the subsection 3.2.2, the holding time for synchronisation may differ from the one that was estimated when making the decision to synchronise, due to the misestimating of vehicle arrival times. It is, therefore, important to examine the extent of this difference. To this end, the distribution of the unexpected holding time for synchronisation is used. This unexpected time can be negative or positive, if the vehicle arrival time prediction was overestimated or underestimated, respectively. A zero value denotes that the vehicle was held as expected by the prediction.

Last but not least, a KPI is defined to assess the impact on vehicle operations. The selection of such an indicator is based on a measure commonly used for determining the fleet size, namely the 90th percentile of the vehicle trip time per line. By holding vehicles at the transfer stop to serve transferring passengers, the time needed for a vehicle to complete its trip is expected to increase. The magnitude of this increase might be a decisive property for public transport authorities and operators regarding the actual implementation of the strategy.

3.3.2 Reference case

The acquisition of values for each of these KPIs does not on its own constitute the evaluation of the system performance. This requires at least one reference case, which provides reference values and allows for a comparison to take place.

The reference case for this study, to be hereafter referred to as 'Benchmark', makes use of the BusMezzo internal controller which holds to restore regularity on the line. The comparison may then highlight the value, if any, of adding the possibility to synchronise transfers.

Apart from this, the results for the four developed controllers are compared to each other, while for each of them, the four levels of real-time passenger data are also distinguished and compared.

3.3.3 Number of runs

Since BusMezzo is a stochastic simulation model, the effect of stochasticity on the results needs to be accounted for. In order to attain statistically robust results, multiple runs are executed. The required sample size to achieve a desired level of reliability is determined using the formula:

$$N' \geq t_{\frac{a}{2}, N-1}^2 \frac{X_s^2}{X_d^2} \qquad (27)$$

where:

N': sample size

$t_{\frac{a}{2}, N-1}$: student-t value for reliability α and a sample N

X_s: standard deviation of the chosen indicator for the sample N

X_d: accepted standard deviation

The phenomenon, whose accuracy and standard deviation are of interest in this case, is the first KPI, namely the perceived trip time from origin to destination per passenger. Since the standard deviation that is exhibited in the population is unknown, an initial set of runs (N) needs to be executed first to determine it. Then, by specifying the desired accuracy, the minimum sample size is computed and applied to all scenarios.

4 Application

In this chapter, the controllers developed in chapter 3 are applied in a case study and their performance is assessed using the BusMezzo simulation model. The selected case study is described in section 4.1. Following the introduction of the case study, a number of scenarios are designed in section 4.2 to demonstrate the value of the controllers and their robustness. The results for each of the scenarios are presented in section 4.3.

4.1 Case study

The simulated network represents two tram lines in The Hague, The Netherlands. Their routes are depicted in Figure 4.1. The operation of vehicles and passengers is simulated only in one direction. Tram line 3 is coloured in blue and the considered direction is the eastbound one, towards Zoetermeer. The red line represents tram line 17 in the southbound direction towards Wateringen. The line configuration and travel times are based on public data provided by the public transport operator in April 2016 [HTM].

More specifically, line 3 consists of 40 stops and line 17 of 35. They intersect in one stop which is marked by a black star in Figure 4.1. This transfer stop is the 15th stop for line 3 and the 7th for line 17 and its location is modelled to be on the same physical point. Since the lines are relatively long, 33 and 16 kilometres, respectively, holding control for regularity has been applied to two additional stops downstream the transfer stop for each line. The choice of positioning them downstream can be justified by the fact that the transfer stop is in the beginning of the lines and there is no stop upstream of high significance. Also, it prevents the insertion of yet another uncertainty in the prediction of the arrival time of vehicles, which would be the case if the vehicles could be held upstream of the transfer stop. The selection of these two stops is made such that they coincide with important transport hubs. For line 3 these stops are #21: Den Haag Centraal and #29: Zoetermeer Centrum-West, while for line 17 they are #17: Den Haag HS and #:28: Rijswijk Station. Based on this, the horizon length, μ, is selected such that it considers the effect on passengers downstream the transfer stop up to the next control point for each line.

In order to ensure that the performance of the real-time controller is assessed rather than the tactical timetable planning, dispatching times are shifted so that the two lines are scheduled to synchronise at the transfer stop. The planned headway for each of the lines is 10 minutes between 7AM and 6PM on weekdays. Adopting the common industry standard, vehicle arrival time predictions are based on the propagation of any existing delay to downstream stops, thereby estimating the arrival time by shifting the scheduled arrival time according to the expected delay [Cats and Loutos, 2016].
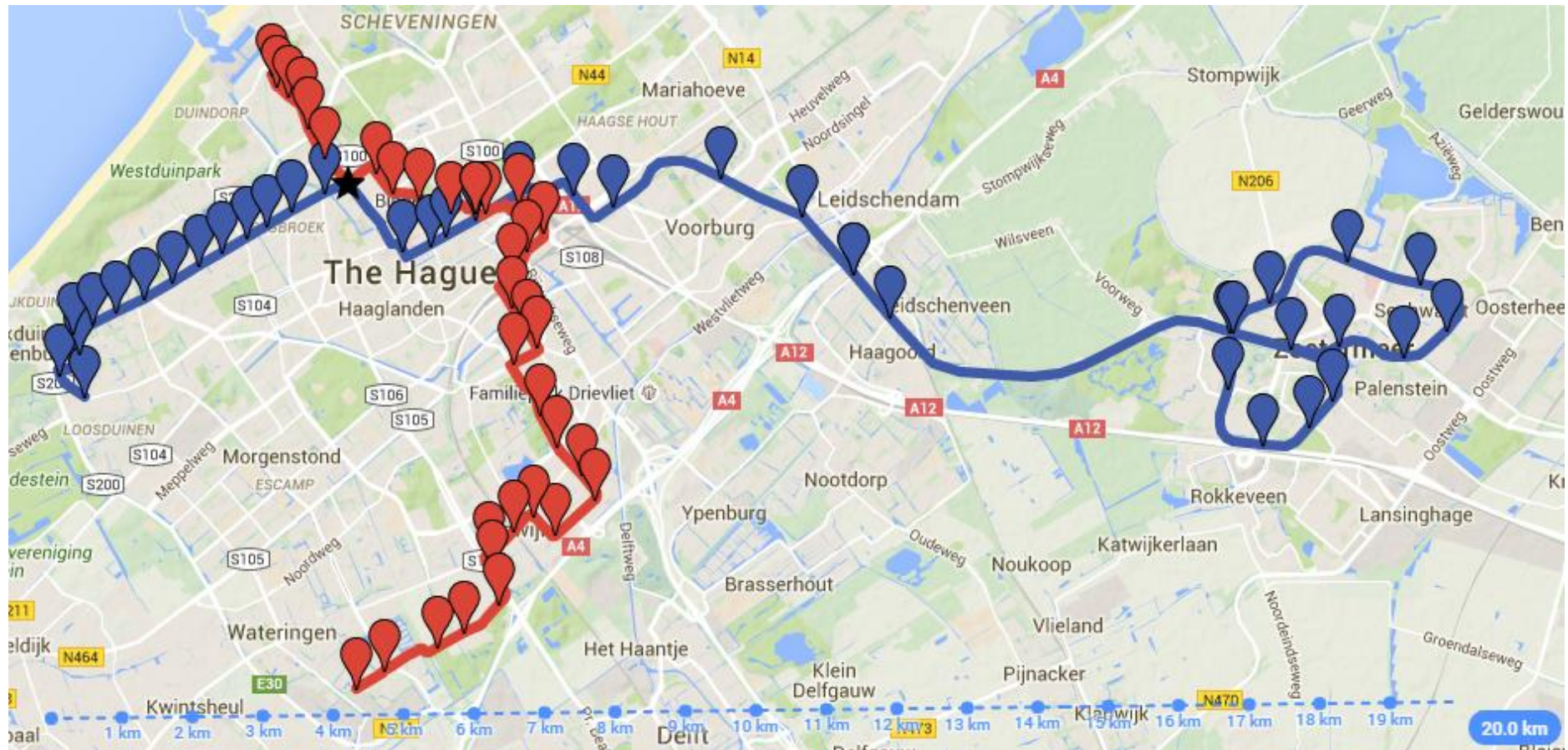
Figure 4.1: Routes of the two simulated tram lines in The Hague.

The dwell time at the stops is determined by the maximum of the boarding and alighting time, since the two processes can take place simultaneously, and it is modelled as such, while also considering a non-linear crowding effect, as suggested by Weidmann [1994].

The tram type operated on line 3 is RegioCitadis with 84 seats and a total capacity of 214 passengers, while GTL8 trams run on line 17. Those have a capacity of 76 seated passengers and a total capacity of 188 passengers.

The demand is defined in terms of OD pairs to allow for the passenger route choice, which however is redundant in the network defined for this case study. Since the designed controller makes use of passenger arrival, alighting and transferring rates for the passenger demand prediction, these are deduced per stop and line and fed into the controller. A visualisation of the OD pairs is given in Figure 4.2. The top, red part represents line 17 and the other line 3. The nodes represent the stops, which are ordered from left to right to follow the direction considered for each line. The links between the stops correspond to the OD pairs that have a positive demand, while the transferring flows are visualised by the single link between the two lines, which is available at the transfer stop. The links from and to the transfer stop are formatted in dashed line style in order to be distinguished from the within-line demand.
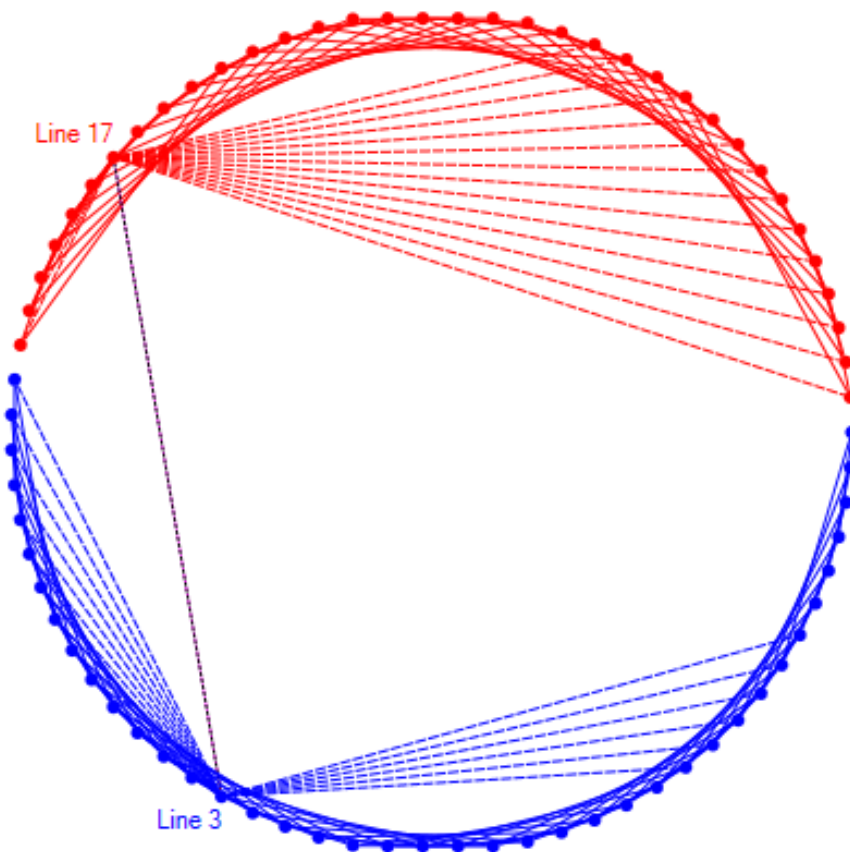
Figure 4.2: Demand representation between OD pairs.

In order to compute the perceived trip time for the passengers (i.e. generalised cost function of Eq. (23)), the weights for each component need to be defined. These are specified based on literature [Cats et al., 2014; Cats et al., 2016; Wardman, 2004] (27; 31; 32): $\{\beta_1, \beta_2, \beta_3, \beta_4\} = \{2, 1.5, 7, 1\}$. These weights are also used for the calculation of the first KPI at the end of the simulation run, when the realised travel time of all passengers is known.

Moreover, crowding multipliers (γ_1 and γ_2 in Eq. (21) and (22)) are applied to account for the perceived in-vehicle riding time as function of the vehicle load factor, the ratio of the vehicle occupancy over the seat capacity based on the findings of Wardman and Whelan [2011]. A visualisation of these multipliers is given in Figure 4.3. The assumption is made that standing passengers appear only after the seat capacity is depleted (i.e. the load factor is greater than 100%). The asterisks correspond to the values found in the study and a stepwise transition between the asterisks has been assumed and implemented in BusMezzo. For consistency, the same stepwise functions have been adopted by the controllers.
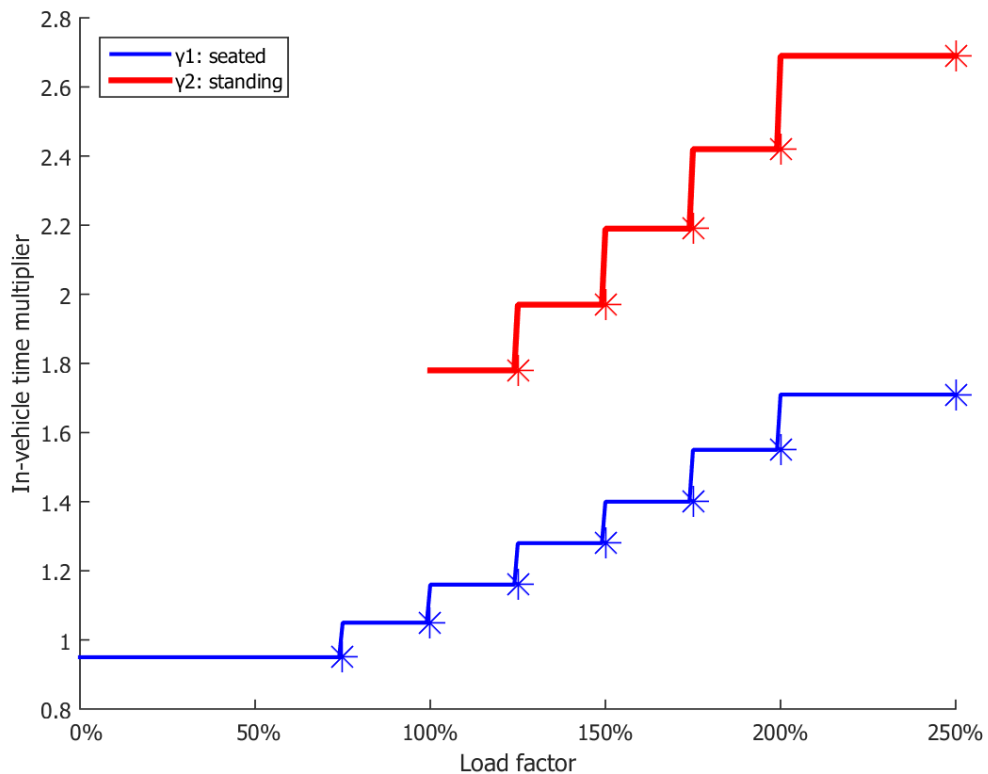
Figure 4.3: Crowding multipliers, adjusted from [Wardman and Whelan, 2011].

4.2 Scenario design

The properties that have been discussed in section 4.1 are used for all scenarios, unless stated otherwise. The property that has yet to be defined is the demand profile, for which three scenarios are designed, expressing a low, a medium and a high vehicle load. The medium load is considered to be the base case scenario, while the other two apply a ±25% change to the passenger volume per group. The hourly vehicle load along the line is displayed in Figure 4.4.
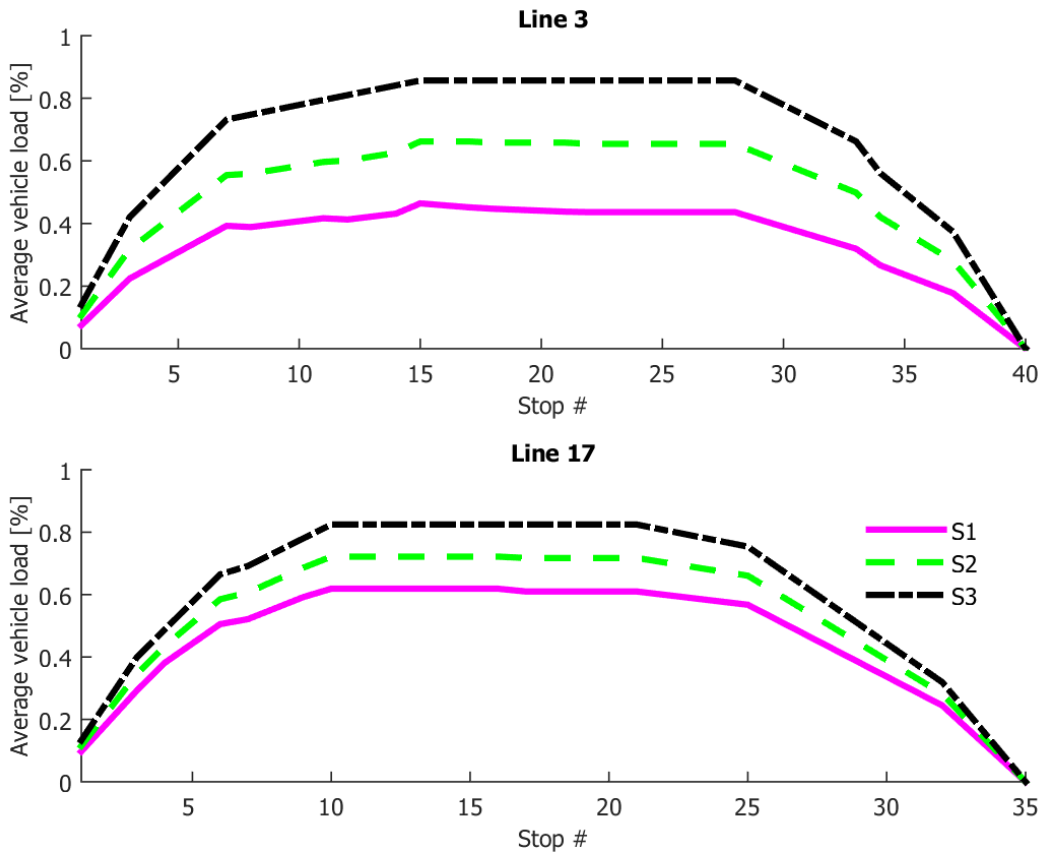
Figure 4.4: Average vehicle load profile per scenario for line 3 (top) and line 17 (bottom).

The fraction of each group is maintained the same in all scenarios and the segmentation of the demand is shown in Table 4.1. Due to the fact that the transfer stop is in the first third of each line, it is reasonable that the highest portion of the passengers originates and is destined to a downstream location. Those transferring are designed to be less than those traversing, which is realistic, since otherwise there would have been a transit line covering the transfer routes instead of the current ones.

Table 4.1: Passenger demand segmentation.

Passenger group	Fraction
Transferring	6%
Traversing	12%
Downstream	62%
Upstream	20%

However, some of the other characteristics of the demand profile change and they are summarised in Table 4.2. As already indicated, the demand level, and thereby the maximum load, are the parameters that are intentionally changed. The maximum load varies by 20% for line 3 among the scenarios and by 10% for line 17. The transferring ratio, which is not a design parameter but rather a result of the design, is slightly different, varying by 2% for the transfer from line 3 to 17, and by 4% from line 17 to 3.

Table 4.2: Scenario properties.

Property\Scenario	1	2	3
Demand level	-25%	Base case	+25%
Maximum load line 3	46%	66%	86%
Maximum load line 17	62%	72%	82%
$\rho\_tr_{3\rightarrow17}$	30%	28%	26%
$\rho\_tr_{17\rightarrow3}$	32%	36%	40%

4.3 Results

In this section, the results of the simulation runs for all scenarios are presented. As already mentioned, a number of runs is executed per case in order to account for the stochasticity in the simulation model and the displayed results average over them. The determination, as well as the evaluation of the sample size used is described in subsection 4.3.1. Based on the scenario design, the reliability of the service can be judged, which is discussed in subsection 4.3.2. Another design parameter for the scenarios has been the horizon length, whose choice is evaluated in subsection 4.3.3. Subsection 4.3.4 presents the share of control decisions that opt for transfer synchronisation. Then, in subsections 4.3.5 – 4.3.7, the results of the individual scenarios are presented.

For consistency, a standardisation has been applied in the naming of the cases displayed in the figures. The title adopts the form of SxCy, where x refers to the scenario number (1: low load, 2: medium load, 3: high load) and y to the controller number (1: PassIT, 2: C^2_PassIT, 3: C^3_PassIT, 4: UC^3_PassIT). Each x-y combination has been simulated for the four different levels of passenger real-time data (hereafter referred to as PRTD). These are: 'None' (use only historic averages), 'Occupancy' (get vehicle occupancy in real-time), 'Tap-ins' (get the demand at the platforms in real-time), 'Both' (combination of Occupancy and Tap-ins). For each scenario, an additional control case has been simulated, where the holding strategy at the transfer stop considers only the regularisation of each line, in order to assess the added-value of real-time transfer synchronisation. This is named 'Benchmark'.

A first set of runs demonstrated that there is a bias in the arrival time prediction of m_2, which was underestimated by one minute in most cases. In order to remove this bias, one minute has been added to the expected arrival time of m_2 for each control decision made by the controllers. The results presented in each section correspond to this corrected prediction. The first result shown is the amount of holding time per line and controlled stop in comparison to the Benchmark in order to see the effect of the control strategy on the lines. Following that, a representation of the KPIs defined in subsection 3.3.1 is displayed:

a) Difference in the perceived trip time from origin to destination per time component and passenger in each passenger group compared to the Benchmark case.
b) Cumulative distribution of the transfer waiting time.
c) Cumulative distribution of the unexpected holding time for synchronisation.

d) Difference in the 90th percentile of the vehicle trip time per line compared to the Benchmark case.

When the results of the control cases with synchronisation are compared to those of the Benchmark, a subtraction takes places which uses the Benchmark as reference. Given this, it is worth keeping in mind that the negative time values observed in the plots correspond to a decreased time and translate into time savings, hence a positive effect.

Before proceeding with the presentation of the results, it is worth commenting on the running time required per decision by each combination of PRTD level, controller and scenario. When there is no uncertainty in the prediction of the demand level, i.e. controllers 1, 2 and 3, the running time per decision is 0.05s regardless of the PRTD level and the scenario, thereby rendering them applicable in real-time.

The introduction of uncertainty in controller 4, given the meshing of 5% in the examined percentile range, leads to higher running times, which are 8s in the None case and 3.5s otherwise. The increased time in the None case is justified by the fact that the predictions need to cover a longer horizon (i.e. from the beginning of the lines instead of the current vehicle locations). These times would have to be reduced before the fourth controller can be applied in real-time, either by changing the meshing or making modifications to the controller.

4.3.1 Sample size

After running 30 replications for the base case scenario (medium demand level), all four controllers and all four PRTD levels using 30 different seeds, the statistics for the perceived trip time per passenger KPI can be computed.

For 30 runs and a reliability of 95%, the student-t value is 2.045 [Student t-Value Calculator]. The desired accuracy is set to $\pm 1.5\%$ of the averaged perceived trip time per passenger. The minimum sample size is computed for each of the 17 cases (4 controllers x 4 PRTD + 1 Benchmark), and their maximum value is defined as the required sample size.

This process designates a minimum of 20 runs, which corresponds to the case with the highest standard deviation (equal to 5.7min), whose averaged perceived trip time per passenger is 176.7min. This result indicates that the use of 30 seeds suffices and therefore, 30 runs are performed for all the scenarios elaborated in section 4.2.

Having determined this sample size, its adequacy needs to be evaluated. This can be done once all simulations have been performed, by looking at the achieved accuracy and comparing it to the desired value of $\pm 1.5\%$. By adjusting the minimum sample size formula to compute the accuracy, it is found that in all cases for these three scenarios, the achieved accuracy complies with the desired value. Therefore, the results presented in the remaining of the chapter correspond to the 30 runs per case.

4.3.2 Service reliability

The first output to be discussed is the service reliability. It can be expressed by the observed vehicle headway at the transfer stop, which is a design parameter for each scenario and should not be greatly affected by the control strategy implemented there. For this reason, it is presented only for the Benchmark case in Figure 4.5. The first three plots show the headway distribution per line and it becomes obvious that for all scenarios line 17 operates 70% of the times at the planned headway of the line (10min).

Contrary to this, the operation of line 3 demonstrates a higher variation around the planned headway, which could be explained by the fact that the transfer stop is after twice as many stops compared to its location for line 17. Moreover, as the demand level increases, the spread of the observed headways becomes even larger and for scenario 3 (high demand) the line deviates from the planned headway in more than 80% of the times.

The fourth plot displays the inter-line arrival headway at the transfer stop, i.e. the time interval between the arrival of a line 3 and a line 17 vehicle. The negative values stem from the definition of this variable, which takes as reference the line of the vehicle that first visited the stop and then computes the observed time arrival differences. Therefore, when the other line arrives first after 10min, a negative value is calculated.

The expectation for this variable is to have a peak around zero and 10min, which is met for the low demand scenario. This expected pattern, however, fades as the demand level gets higher and even shifts to two different peaks around ±5min for scenario 3, which shows that the service becomes unreliable at high demands.
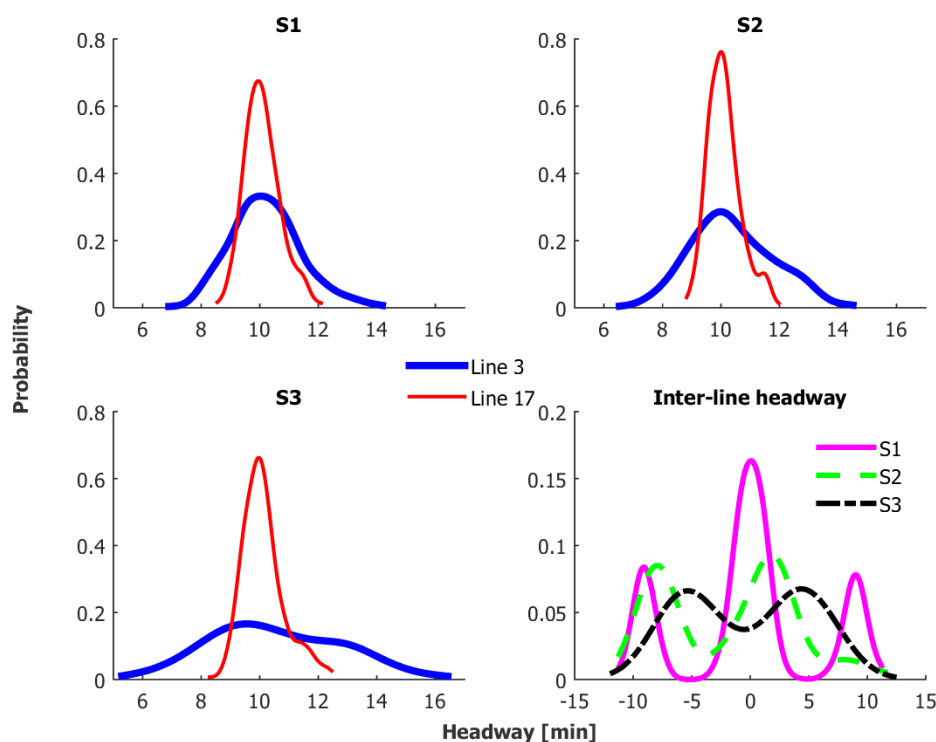
Figure 4.5: Vehicle arrival headway at the transfer stop per line and scenario.

4.3.3 Choice of horizon length

Figure 4.6 illustrates the effect of the control strategy at each of the downstream stops. The zero in the horizontal axis corresponds to the transfer stop itself. The effect is calculated by averaging the perceived trip time for passengers originating at each downstream stop over the four controllers and the four PRTD levels and comparing it to the corresponding value in the Benchmark case.

This analysis could indicate the horizon length that should be used for each line and scenario, since the more the downstream passengers are affected by the control strategy, the more they should be included in the horizon considered in the decision making process. Arguably, this outcome is dependent on the horizon that was used to retrieve it, but it can be viewed as a direction.

A general finding, whose legitimacy is demonstrated in the presentation of the results per scenario, is that line 17 is the one that makes the transfer synchronisation decisions. Based on this, it is expected, and proven to be so, that the effect on passengers of line 3 is minor, even favourable. Therefore, the choice of a horizon length of 6 stops is considered to be adequate.

The situation is different for line 17, where the effect is unfavourable for passengers originating at the transfer stop and diminishes towards downstream locations. The selected horizon length of 10 stops seems to suffice for the low demand scenario, while it is possible that a higher value would have yielded better performance for the other two scenarios. Nonetheless, the results to be presented in the remainder stick to the choice of the 10 stops, which has been the original design, justified by the fact that after 10 stops there is a new control point, where the effect on downstream passengers is going to be evaluated again in the process of making a new control decision.

4.3.4 Transfer synchronisation

The share of synchronisation control decisions made by the developed controllers for each scenario under the different passenger data levels is depicted in Figure 4.7. Given the 10min headway, if the number of vehicles that visit the transfer stop from each line within the simulation time is equal and the reliability is high, the maximum share to be expected is 50%. This would indicate that the first vehicle that arrives will always wait for the upcoming one.

In the low demand scenario, this maximum share can be expected since the service is reliable, even though the amount of transferring passengers may be small. The availability of real-time passenger data leads to this share for all controllers, while controllers C3 and C4 are more conservative in the None case, since only 15% of the vehicles synchronise. The inclusion of the riding time component in these controllers affects the generalised cost estimation and in the absence of real-time passenger data they estimate that it is better for the passenger level of comfort if those transferring board the next vehicle.
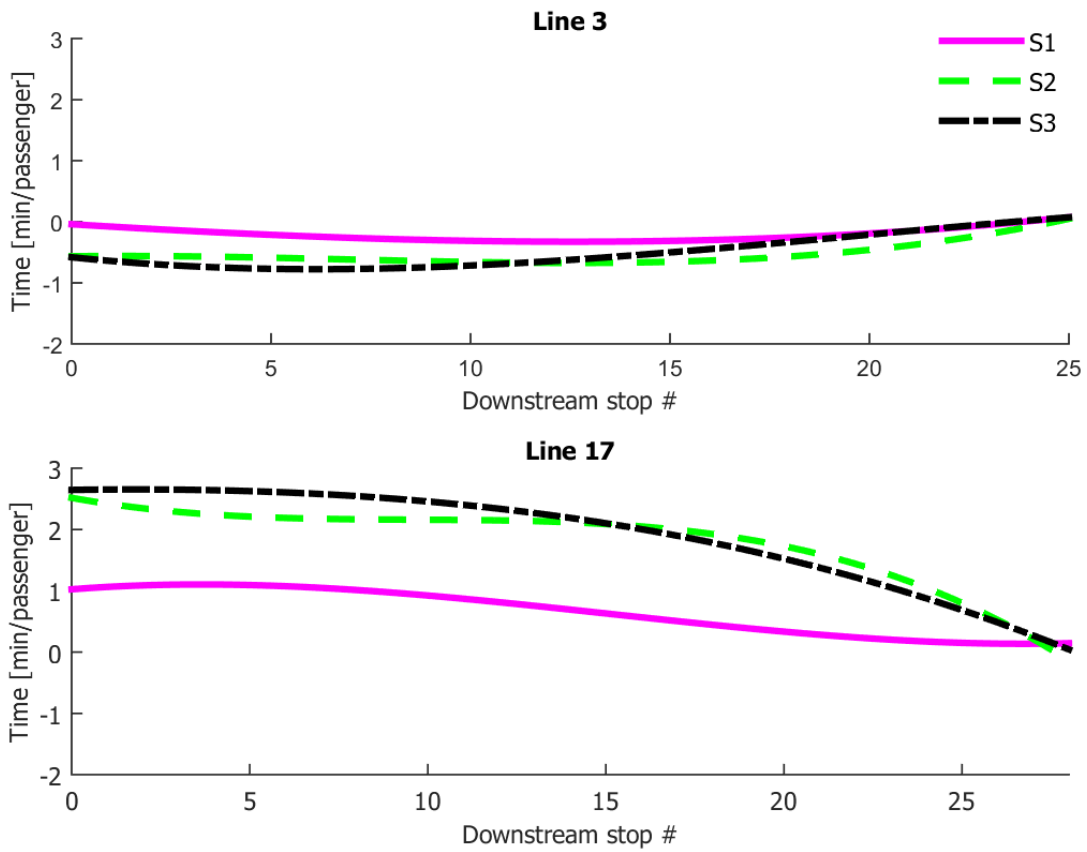
Figure 4.6: Difference from the Benchmark case in the perceived total time for downstream passengers per stop, line and scenario, averaged over all controllers and passenger data levels.
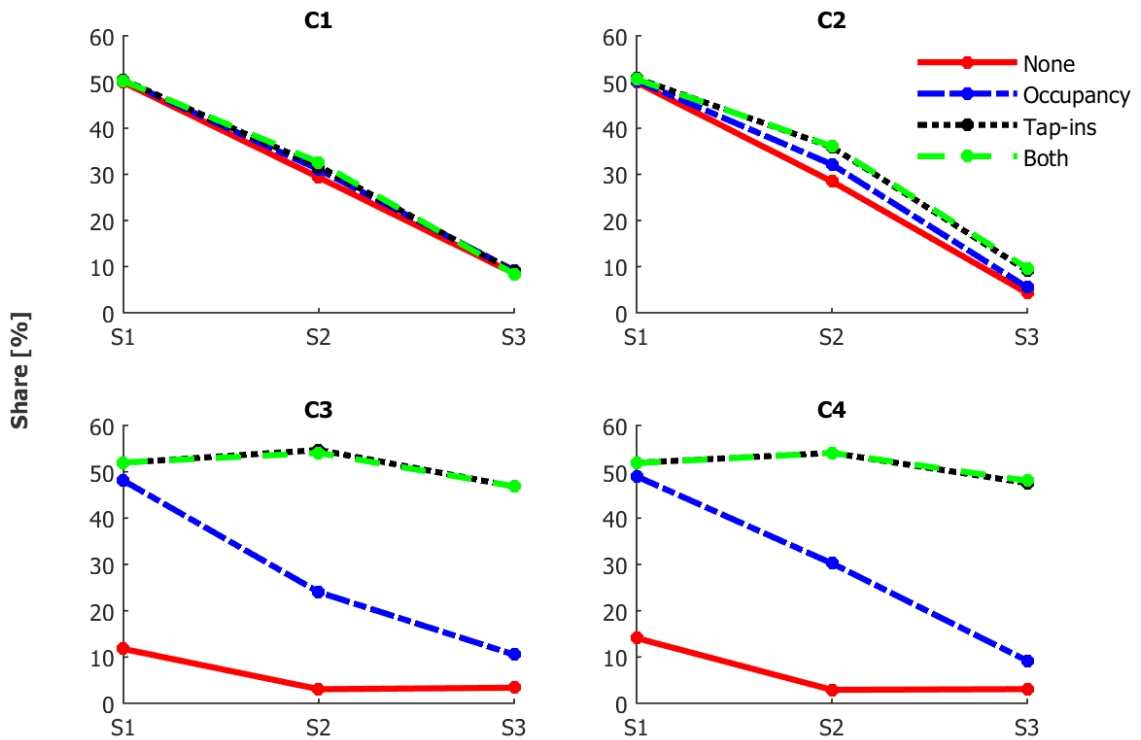

Figure 4.7: Share of transfer synchronisation control decisions per scenario, controller and passenger data level.

As the demand increases, less vehicles are synchronised by controllers C1 and C2, regardless of the passenger data level. In contrast, C3 and C4 follow this pattern only in the Occupancy case. This can be explained by the fact that in this case, the on-board crowding is known in real-time and needs to be predicted only for downstream locations. The acquisition of tap-ins, with or without occupancy data, becomes leading in the decision and the controllers opt in favour of synchronising all transfers. This is attributed to the myopic view of the network load and the ignorance of the effect of the reduced service reliability.

4.3.5 Scenario 1: Low demand profile

Figure 4.8 depicts the average holding time per line and controlled stop in comparison to the Benchmark case. It can be seen that in all cases line 17 is the one that is held at the transfer stop for about one minute, which serves the transfer synchronisation. Line 3 is rarely held at the transfer stop, while there is almost no extra holding time at the other stops. That is because, by holding to synchronise, the line regularity is not disturbed in this low demand scenario and there is no need for correcting actions at the downstream controlled stops.
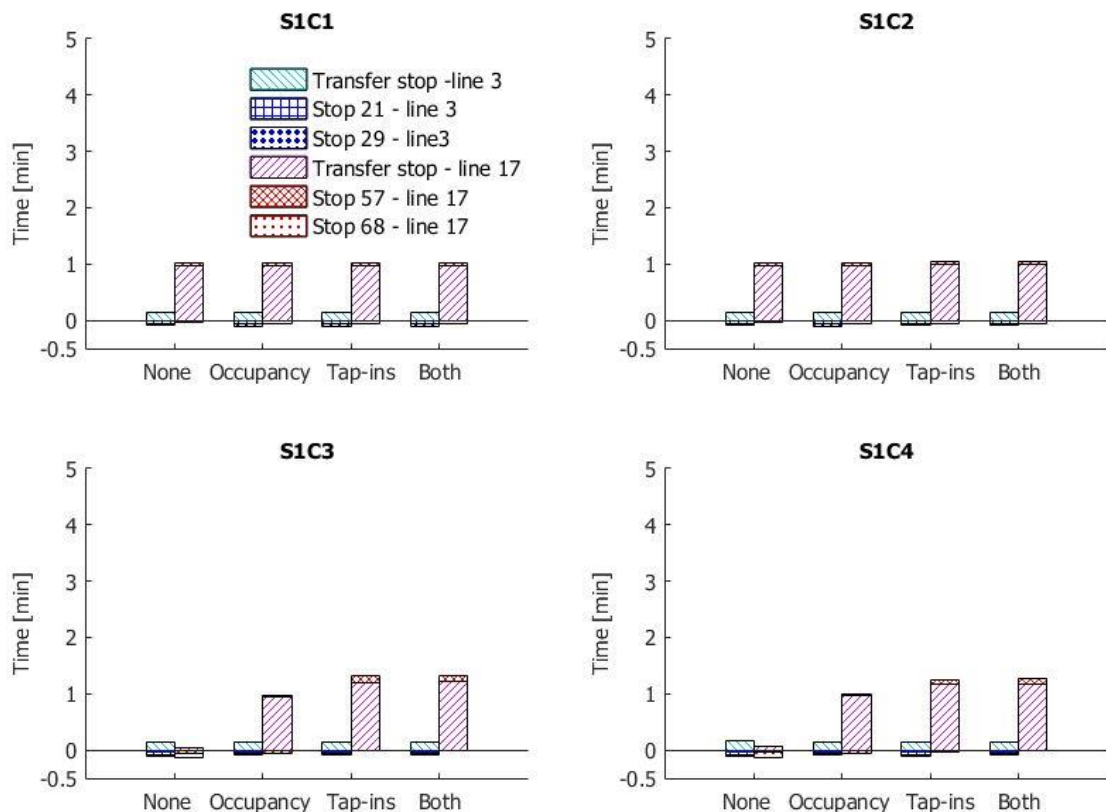
Figure 4.8: Difference compared to the Benchmark case in the average holding time per line and controlled stop for scenario 1.

In Figure 4.9 to Figure 4.12, the difference in the perceived trip time from origin to destination per passenger is visualised for the four designed controllers. The difference is expressed compared to the Benchmark case for scenario 1, while distinguishing the six time components and the four passenger groups. A number of remarks can be made by examining these figures.

To begin with, it can be seen that there is no effect on upstream passengers in all cases. That is reasonable, since they board and alight before any control decision is made, and, therefore, are indifferent to the control strategy that is applied at the transfer stop. In this scenario, a great benefit (up to 10min) is attainable per transferring passenger. Their waiting time at the transfer stop is significantly reduced, while their riding time is also improved, indicating that they are more evenly spread among vehicles. Moreover, the effect on downstream passengers is only a slight increase in their riding time, whose magnitude can be considered negligible. The travelling experience for the traversing group deteriorates, since they are held at the transfer stop for about 1.5min per passenger and their riding time increases by a small amount because they have to share the vehicle with the transferring passengers.

The outcomes of the controllers without on-board crowding consideration (i.e. C1 and C2) demonstrate similar performance with respect to this KPI, among each other as well as among the four PRTD levels. This is reasonable given that their only difference is the inclusion of capacity constraints in C2, which remain inactive in this low demand scenario. The other two controllers exhibit a similar behaviour when real-time passenger data is available. In the None case, however, the benefit for transferring passengers is much lower (4min/passenger) and the effect on the other groups can be considered negligible. This result relates to the share of synchronisation decisions, which are less frequent in this case.
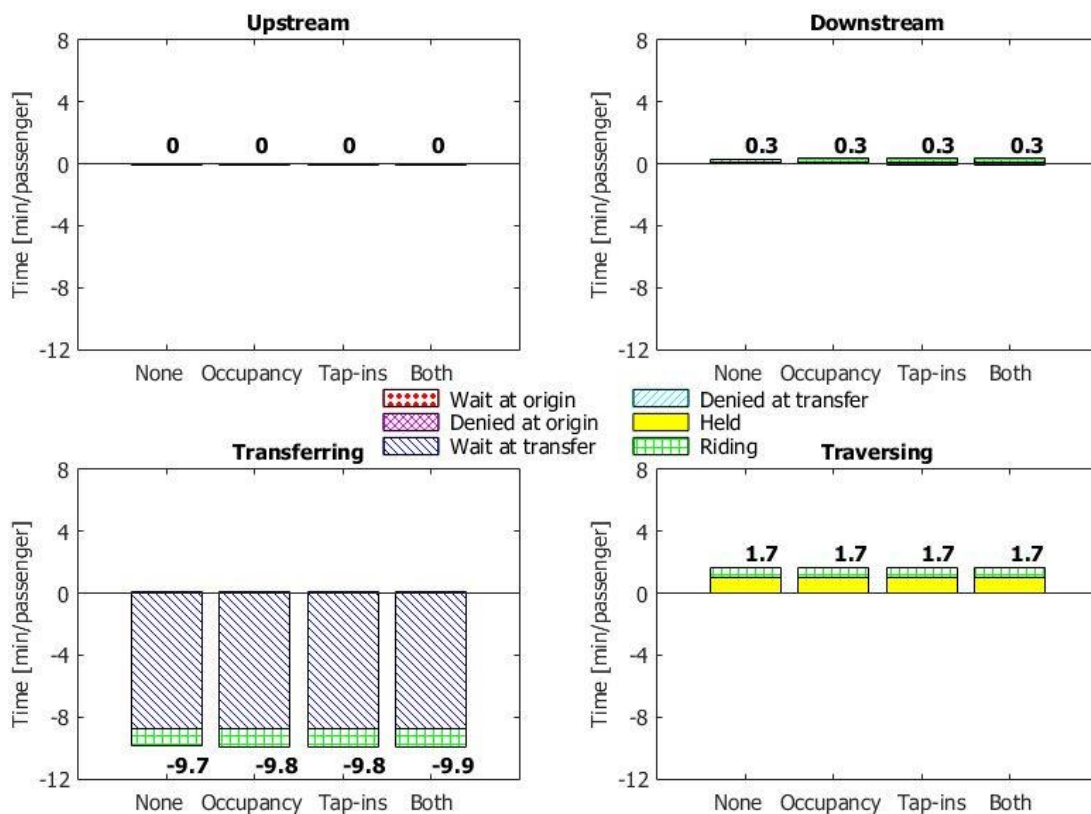

Figure 4.9: Difference compared to the Benchmark case in the perceived trip time from origin to destination per passenger in each passenger group for the case S1C1.
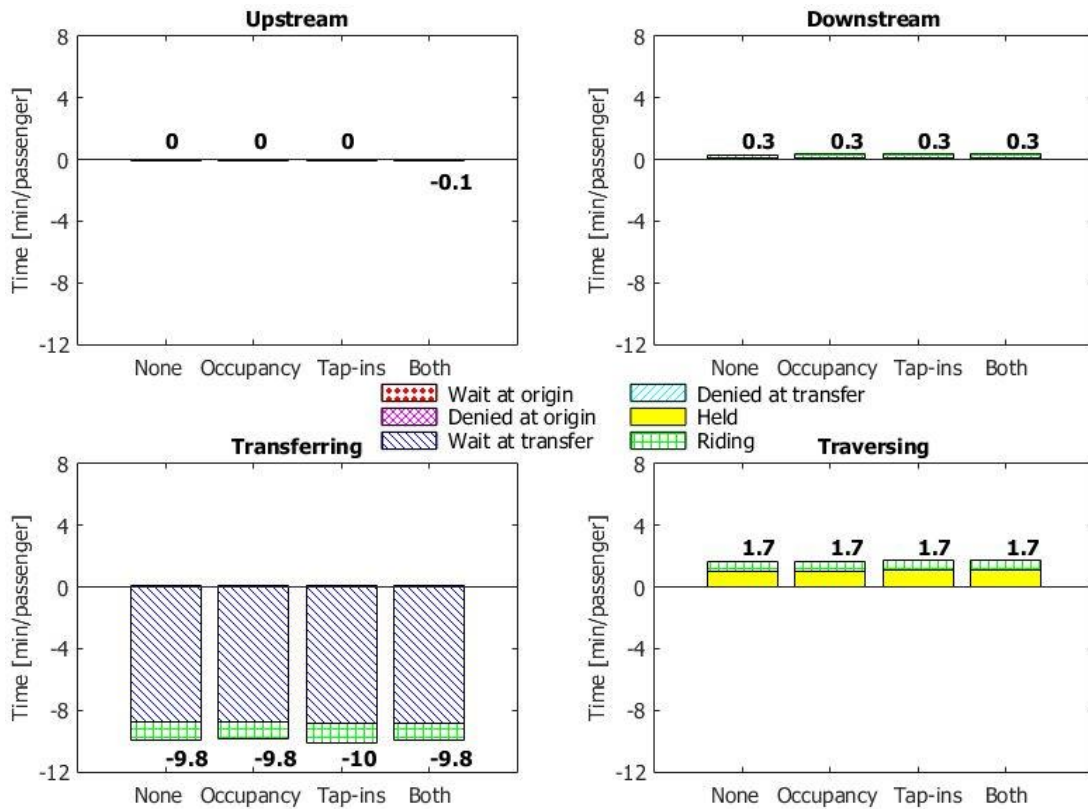
Figure 4.10: Difference compared to the Benchmark case in the perceived trip time from origin to destination per passenger in each passenger group for the case S1C2.
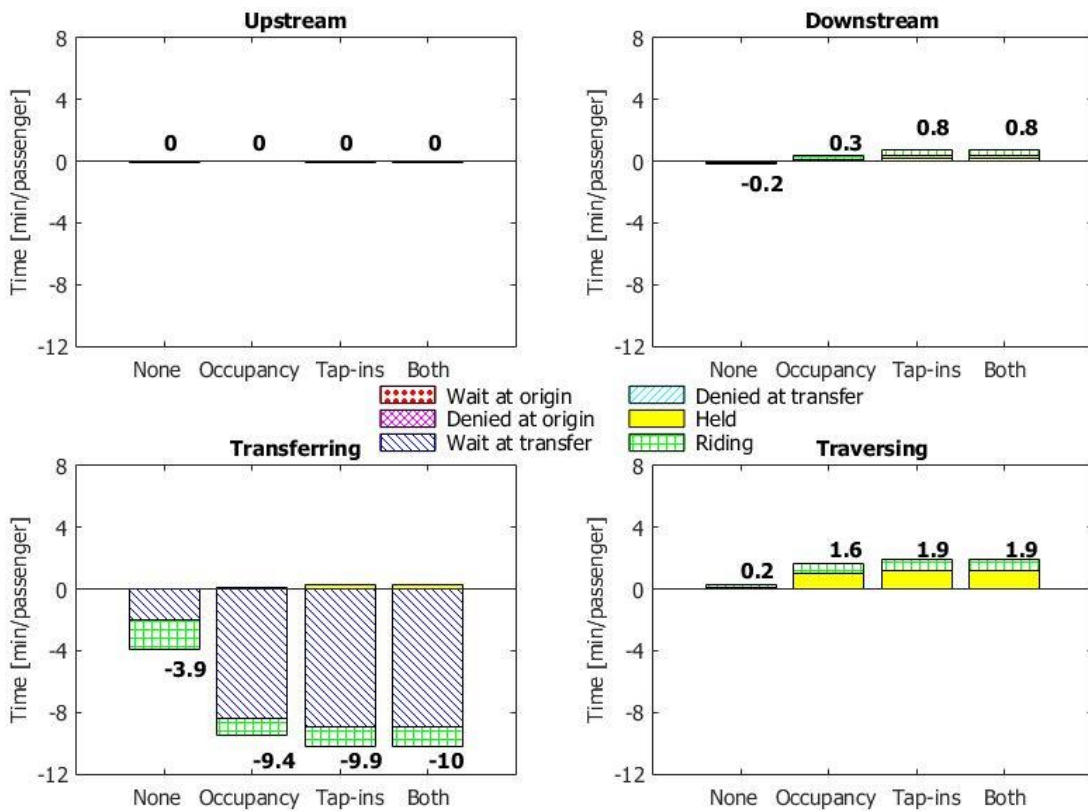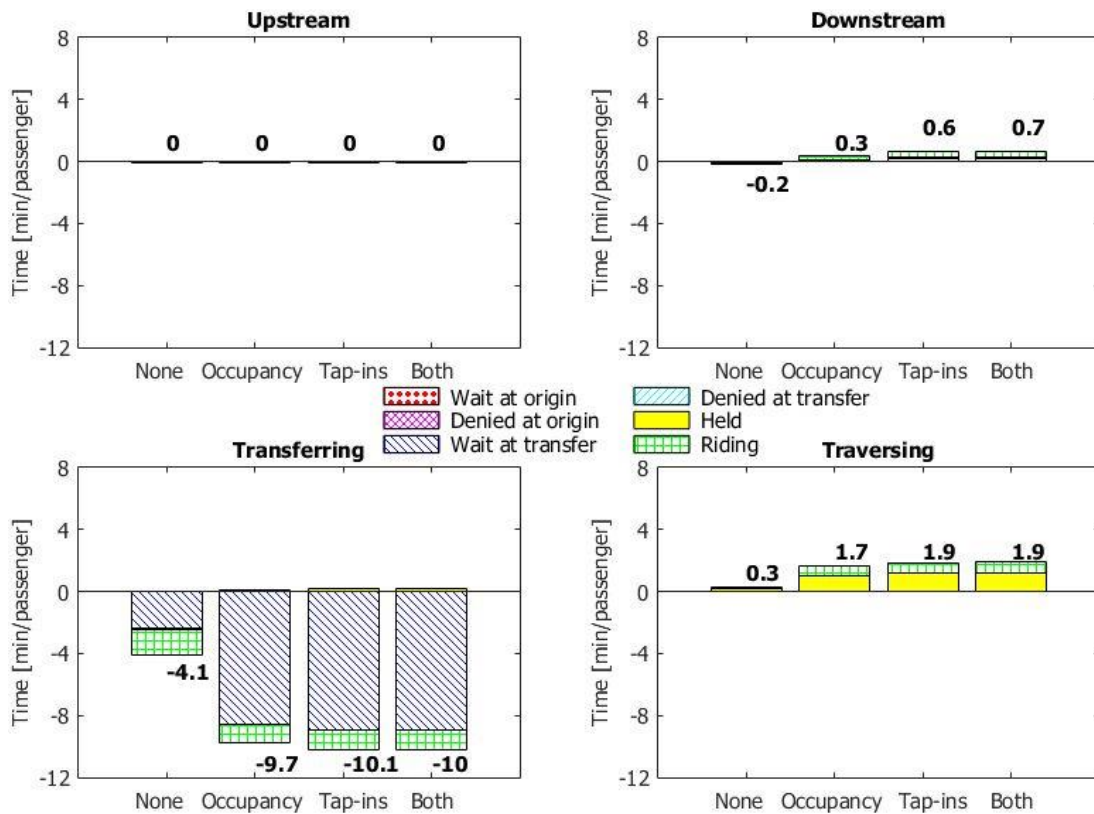

Figure 4.11: Difference compared to the Benchmark case in the perceived trip time from origin to destination per passenger in each passenger group for the case S1C3.

Figure 4.12: Difference compared to the Benchmark case in the perceived trip time from origin to destination per passenger in each passenger group for the case S1C4.

The time savings for transferring passengers are further investigated by examining the cumulative distribution of the transfer waiting time depicted in Figure 4.13 per passenger data level and controller for scenario 1. The line corresponding to the benchmark is the same in all graphs since it depends only on the scenario (i.e. demand profile) under examination.

The shape of the curves represents the headway between successive arrivals of vehicles from different lines at the stop. In the Benchmark case, there is a step between 2.5 and 8 minutes at 50% probability. This clearly distinguishes two classes of transferring passengers, those who either transfer directly or alight the vehicle controlled for synchronisation and need to wait for the connecting one to arrive, and those who have just missed their connection and have to wait a full headway.

In the controlled cases, these curves shift upward due to the synchronisation of transfers. Passengers from the second group experience a direct transfer, i.e. wait 0 minutes, and depending on the share of synchronisation, the corresponding part of this group is served, changing the shape of the curve accordingly. The probability to attain a direct transfer is 7% in the Benchmark case and it rises to 25% when controllers C1 and C2 are used and when real-time data is available for C3 and C4.

The difference among the passenger data levels is mostly visible for C3 and C4, where no real-time data leads to results close to the Benchmark, the knowledge of the occupancy results

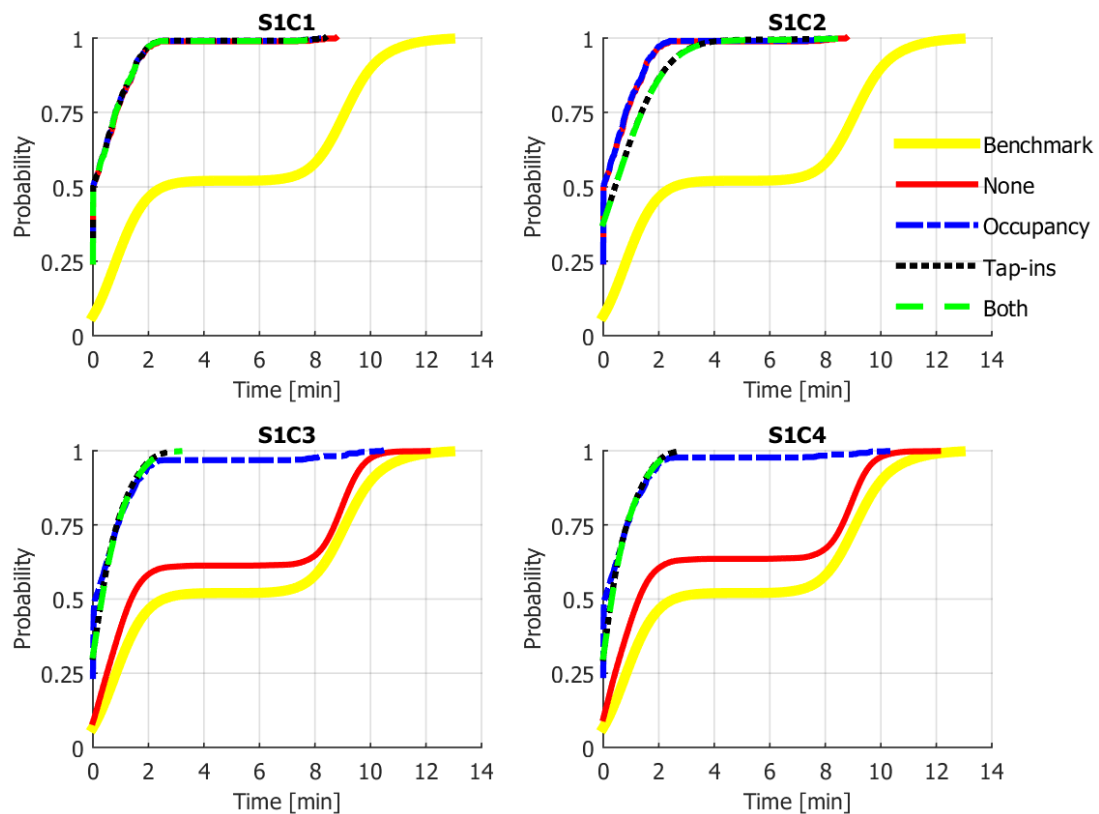in a similar performance to C1 and C2, and the availability of tap-ins causes all vehicles to synchronise.

Figure 4.13: Cumulative distribution of the transfer waiting time for scenario 1.

In Figure 4.14 the cumulative distribution of the unexpected holding time for synchronisation is plotted for each controller. A first remark out of this figure is that there is a 50% probability to have the correct estimate for the expected holding time (i.e. 0 unexpected minutes), which proves that the correction applied in the arrival time prediction of the connecting vehicle has removed the bias. When misestimated, the actual holding time is off by up to one minute in most cases.

Figure 4.15 presents the difference in the 90th percentile of the vehicle trip time per line compared to the Benchmark case for scenario 1. The first observation to be made is that the effect on the two lines is different; while the trip time for vehicles of line 17 increases, it decreases for vehicles of line 3. As already shown in Figure 4.8, line 17 is the one that makes most of the transfer synchronisation decisions. Consequently, the 90th percentile of the trip time of line 17 vehicles is prolonged by up to 2min, with the average holding time being 1min. Considering that the total trip time for each line is around one hour, this may be an acceptable deviation from the initially desired operation. Therefore, it may be found to be negligible, as is the case for the reduction observed for line 3.
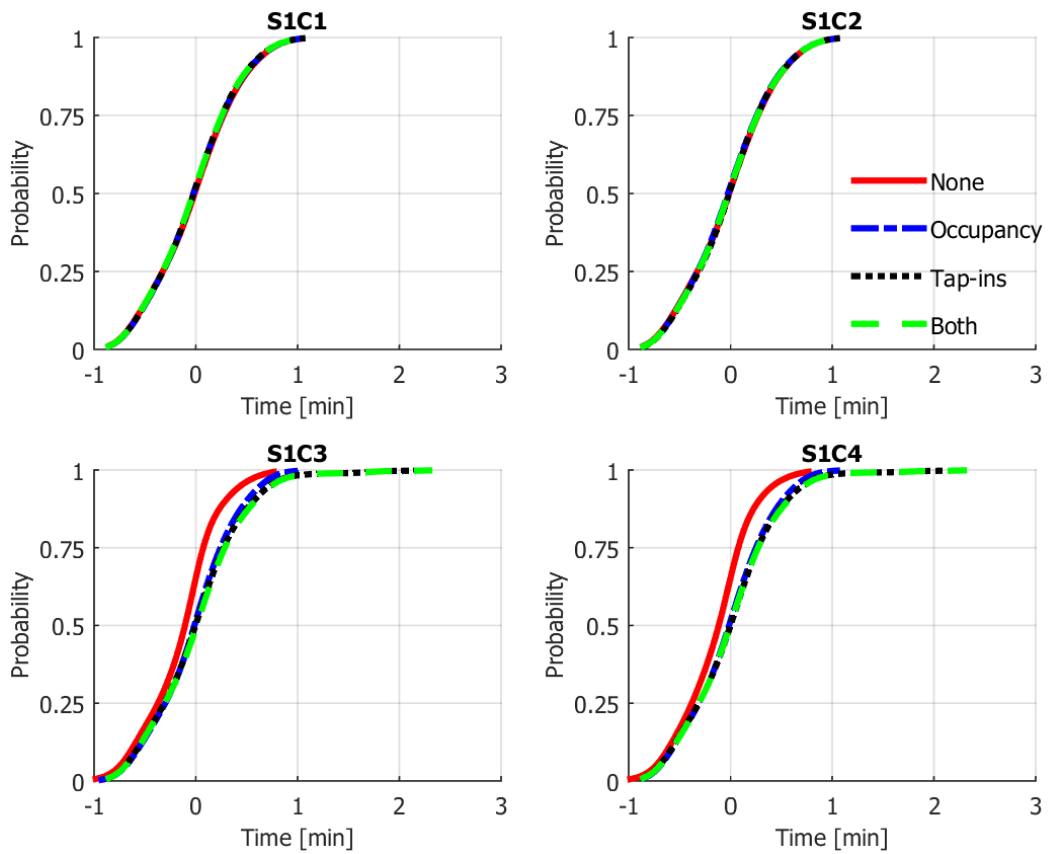
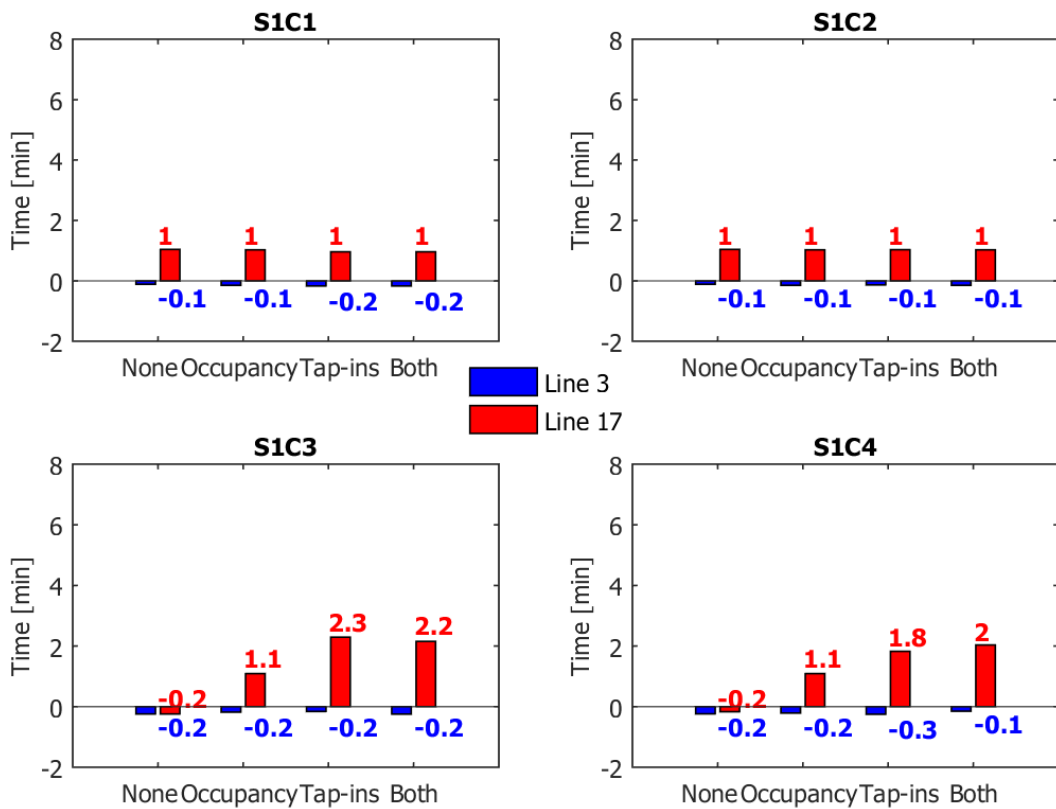Figure 4.14: Cumulative distribution of the unexpected holding time required for synchronisation in scenario 1.

Figure 4.15: Difference in the 90th percentile of the vehicle trip time per line for scenario 1 compared to the Benchmark case.

4.3.6 Scenario 2: Medium demand profile

Figure 4.16 shows the average holding time per line and controlled stop in comparison to the Benchmark case for scenario 2. Line 17 is again the one that is held for synchronisation since it is the one arriving with a shorter interval between vehicles of different lines at the transfer stop and, therefore, more likely to be synchronised. The holding time for synchronisation is between 1 and 3 minutes, which necessitates holding to restore regularity in line 17 at downstream stops. When the holding time at the transfer stop is 1min then a small correction at the first downstream stop suffices, while in case it is 3min the vehicles of line 17 need to be held at both downstream stops. Line 3 is rarely held at any stop.
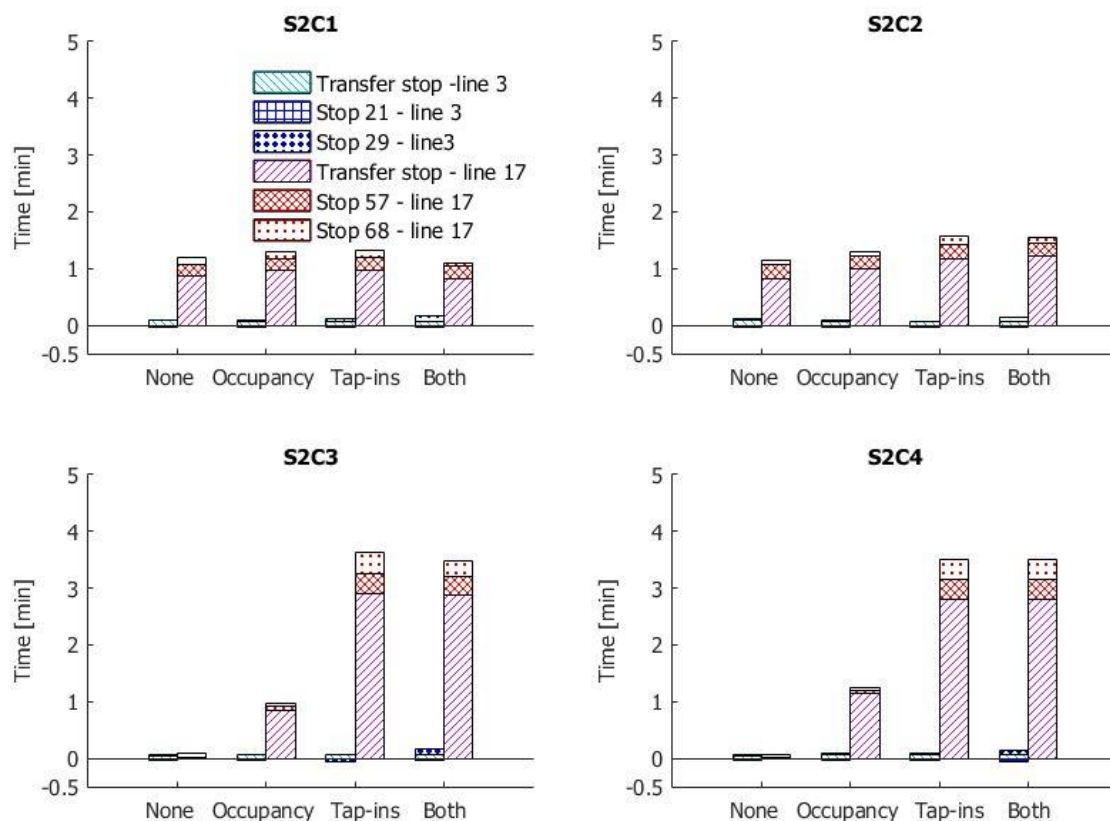

Figure 4.16: Difference compared to the Benchmark case in the average holding time per line and controlled stop for scenario 2.

In Figure 4.17 to Figure 4.20, the difference in the perceived trip time from origin to destination per passenger is visualised for the four designed controllers. The first two controllers achieve great savings in the transfer waiting time, yet they do not consider the on-board crowding and their control decisions lead to uneven vehicle loads and an increase in the riding time of the transferring passengers. As a result, their saving amounts to less than 2min per passenger, which is also the time each traversing passenger experiences as a consequence of the strategy. Given the fact that the fraction of traversing passengers is greater than that of transferring ones, and the small delay incurred by downstream passengers, it can be stated that these two controllers perform poorly under this scenario.

In contrast, the other two controllers take the on-board crowding into account and their performance depends on the PRTD level. When no real-time passenger data is available, the controller rarely synchronises, while the availability of tap-ins leads to the highest share of synchronisation decisions and the greatest benefit for transferring passengers. However, as previously mentioned, the service regularity is disturbed and all passengers experience a holding time at downstream stops. The waiting time for downstream is even increased in this case, which, considering their relative volume, compromises the overall benefit. The results obtained in the Occupancy case are better for the system. The transferring passengers save up to 5min each, while there is no effect on downstream passengers and those traversing experience a delay of 2min each. Consequently, the vehicle occupancy is concluded to be the most valuable real-time data source in this scenario.
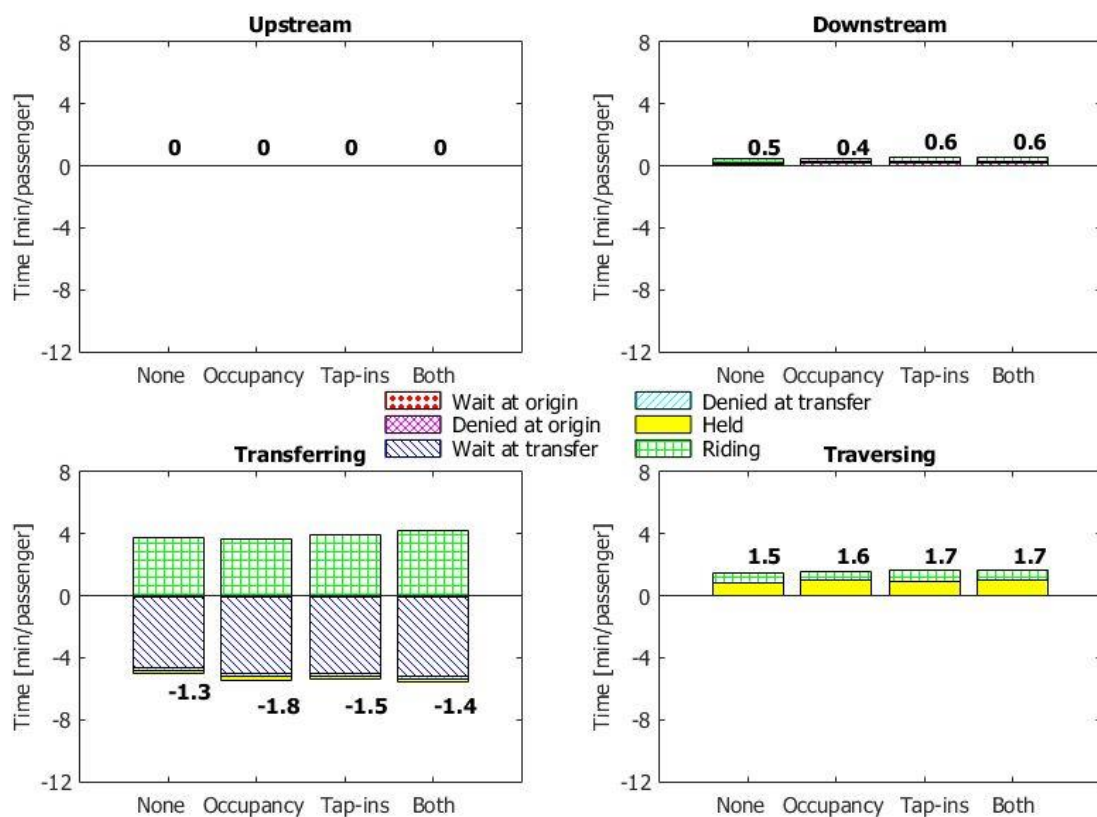

Figure 4.17: Difference compared to the Benchmark case in the perceived trip time from origin to destination per passenger in each passenger group for the case S2C1.
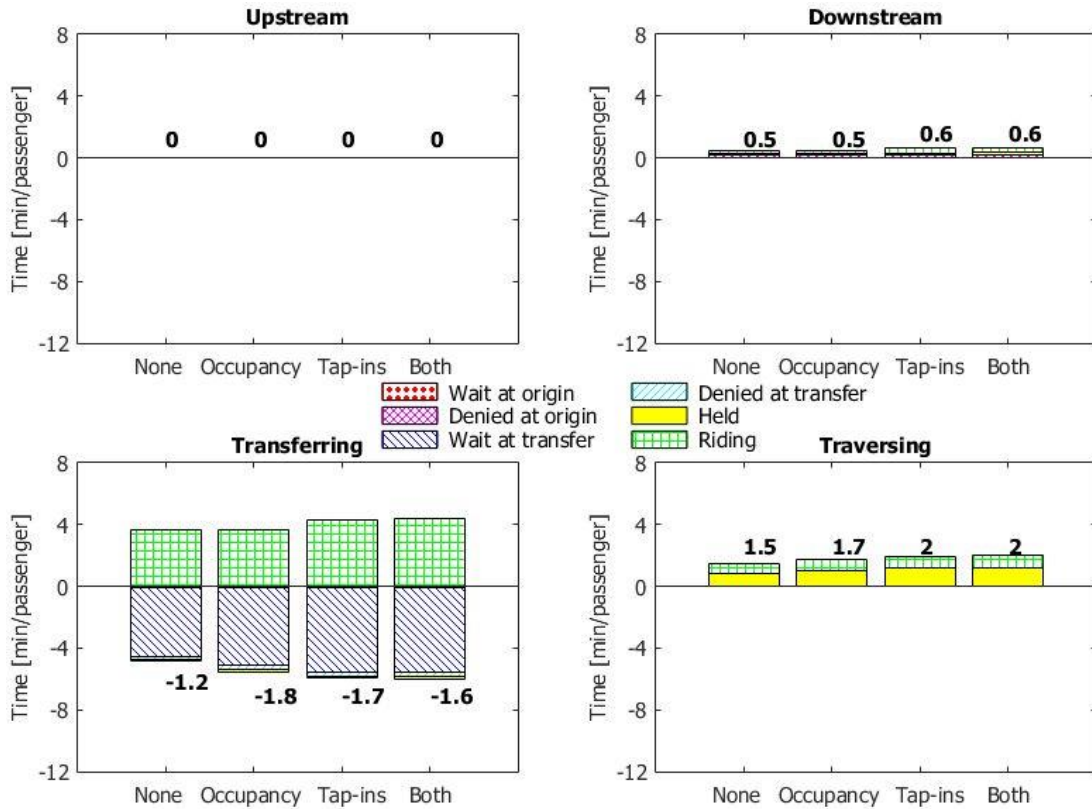
Figure 4.18: Difference compared to the Benchmark case in the perceived trip time from origin to destination per passenger in each passenger group for the case S2C2.
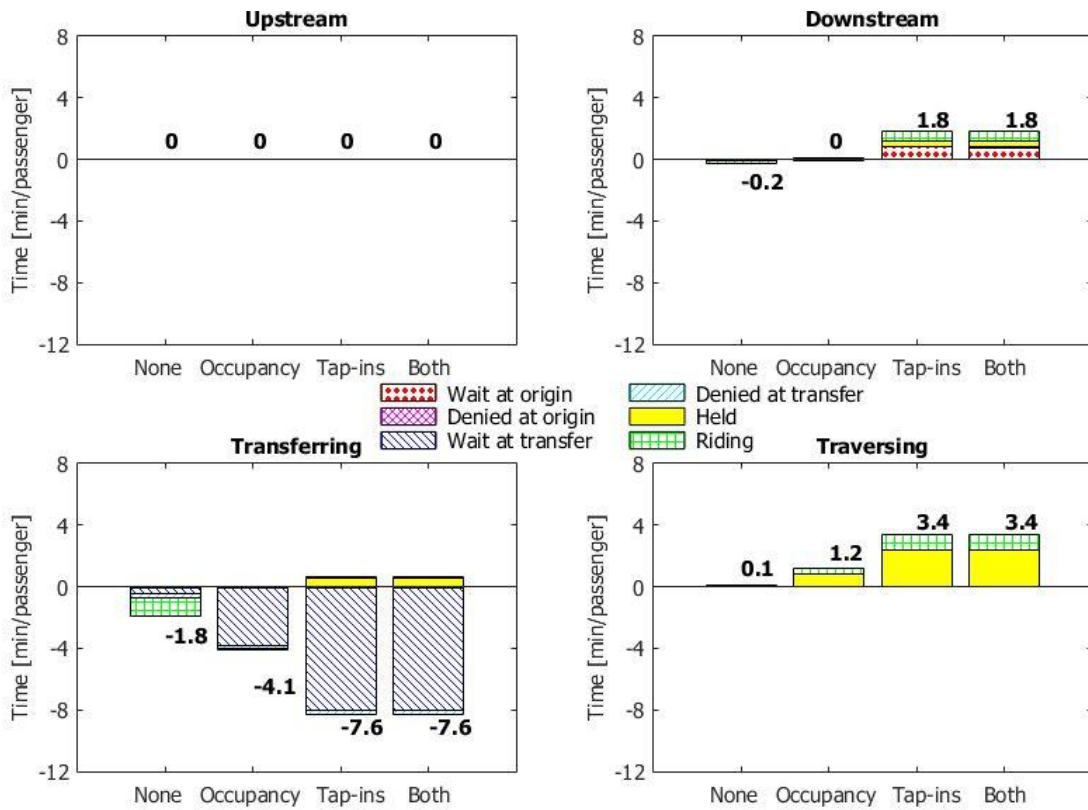

Figure 4.19: Difference compared to the Benchmark case in the perceived trip time from origin to destination per passenger in each passenger group for the case S2C3.
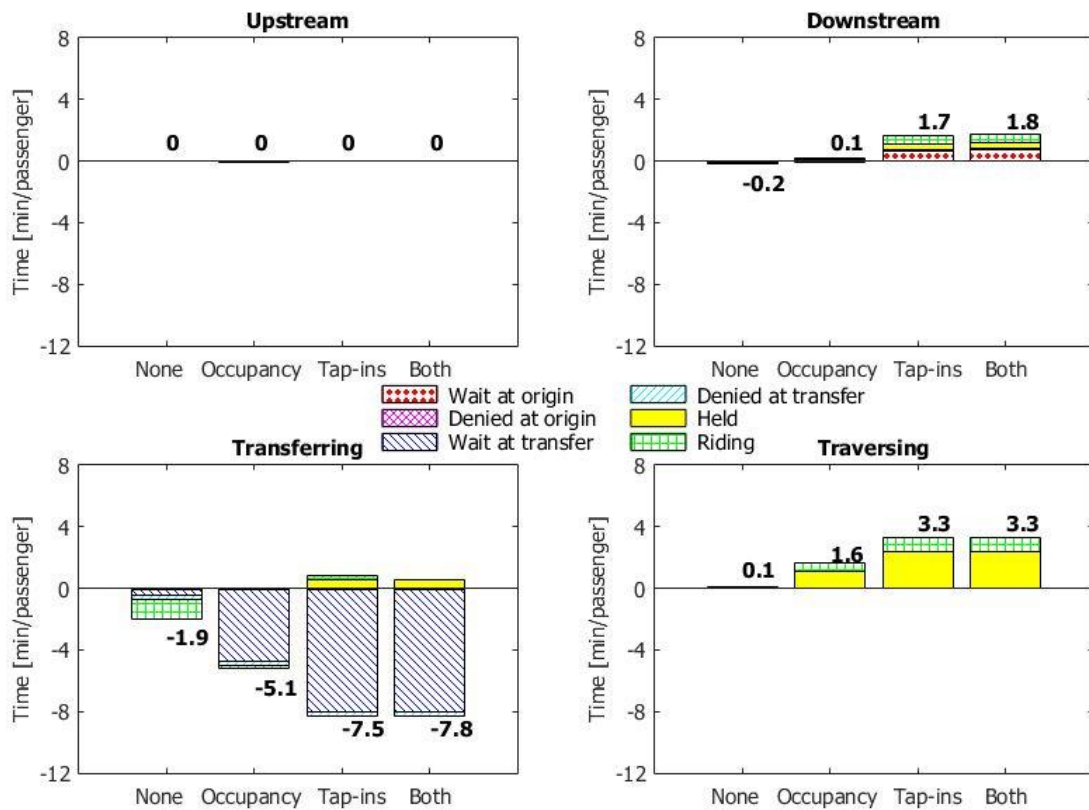
Figure 4.20: Difference compared to the Benchmark case in the perceived trip time from origin to destination per passenger in each passenger group for the case S2C4.

Figure 4.21 shows the cumulative distribution of the transfer waiting time per controller and PRTD level. In the Benchmark case, a small fraction of the transferring passengers waits for more than 20min to complete the transfer, which, given that the vehicle headways is 10min, means that they have been denied boarding. This is prevented in the controlled cases, which can also be observed as a small saving in the perceived time for transferring passengers in the previous figures. Moreover, it can be seen that 50% of the transferring passengers need to wait up to 4min in the Benchmark case, while the same percentage is served in 2min by the first two controllers.

The difference between the PRTD levels for controllers C3 and C4 is visible also in the transfer waiting time distribution. When no real-time data is available, a small amount of transfers are synchronised, which prevent the transfer waiting time due to denied boarding. The Occupancy case performs similarly to the other controllers, with 15% of the passengers transferring directly. When tap-ins are available, all vehicles opt to hold for synchronisation and 98% of the transfers take place within 4min.
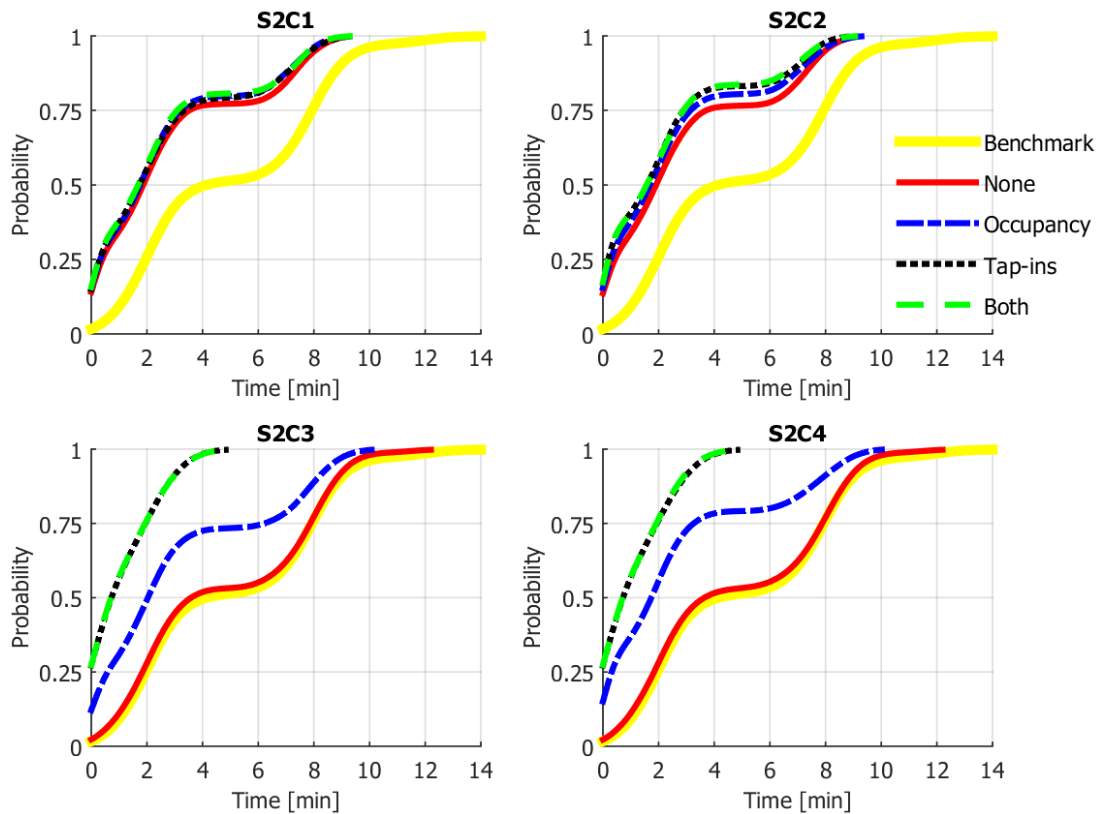
Figure 4.21: Cumulative distribution of the transfer waiting time for scenario 2.

In Figure 4.22 the cumulative distribution of the unexpected holding time for synchronisation is plotted for each controller. Due to the deterioration of the service reliability, the vehicle arrival time predictions become less reliable and the required holding time is in some cases underestimated by up to 3min. This occurs when all vehicles are requested to synchronise and the assumption regarding fully reliable arrival time predictions is bound to be violated.

Figure 4.23 displays the difference in the 90th percentile of the vehicle trip time per line compared to the Benchmark case. In this scenario, the 90th percentile of the trip time of line 17 vehicles is prolonged by as much as 6.7min in the cases when tap-ins are available to controllers C3 and C4 and half of this time is the average holding time for synchronisation. The other half stems from an increase in the dwell time at downstream stops, where more passengers are generated while the vehicle is held at the transfer stop. In the other, less extreme, cases, the additional time for line 17 is in the order of 2.5min, while the saving for line 3 amounts within 0.5 and 1min. This reduction denotes that the service of line 3 becomes slightly more regular. Given the fact that the dwell time is a function of the on-board crowding, a possible explanation for this could be that a more even demand distribution among the line 3 vehicles is achieved, which in turn translates into decreased dwell times at the stops.
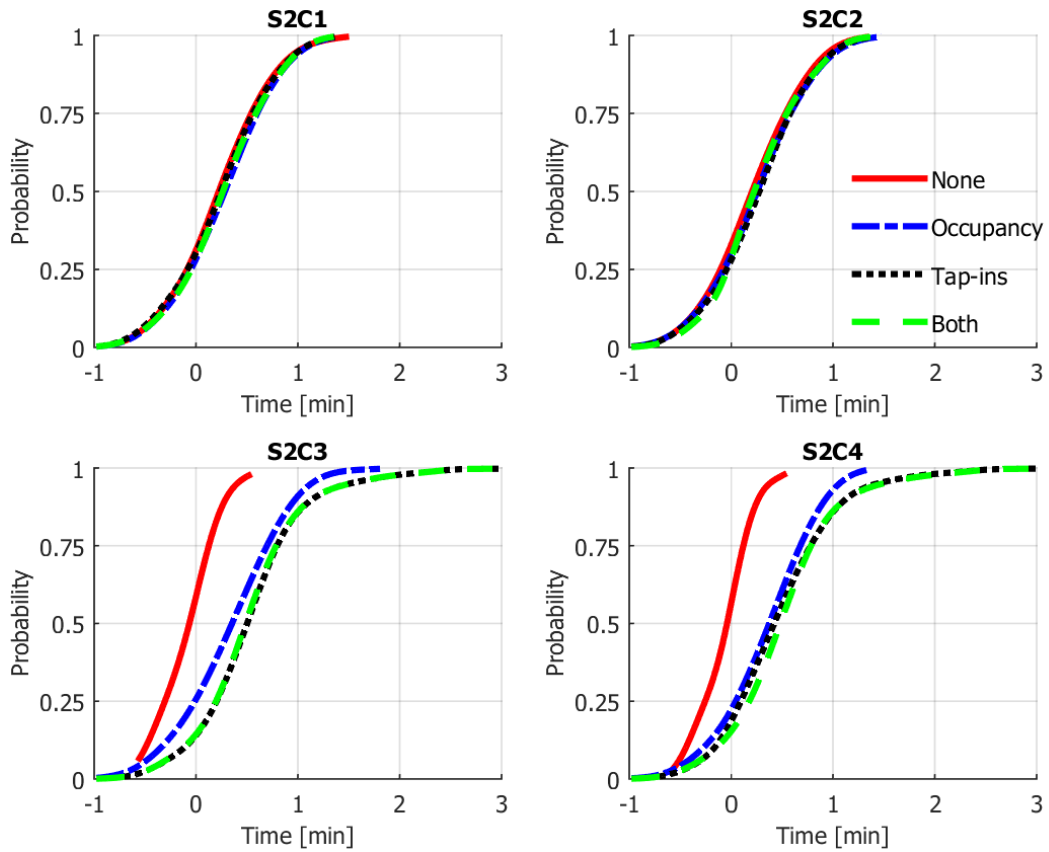
Figure 4.22: Cumulative distribution of the unexpected holding time for synchronisation for scenario 2.
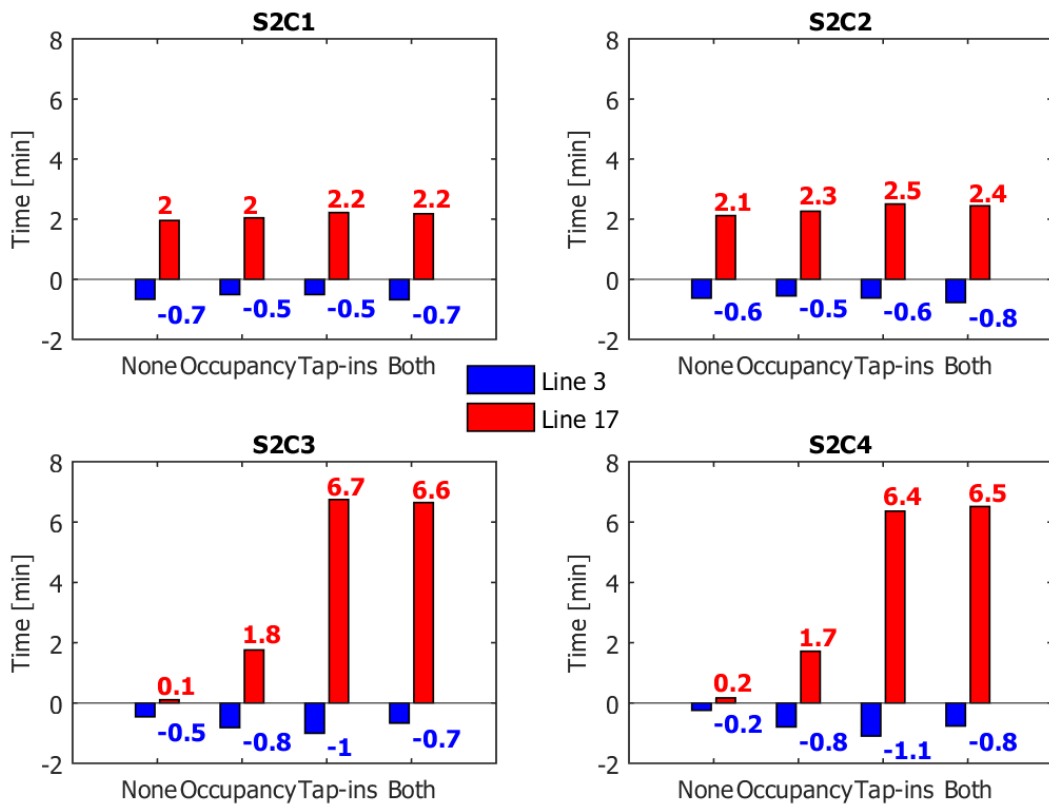

Figure 4.23: Difference in the 90th percentile of the vehicle trip time per line for scenario 2 compared to the Benchmark case.

4.3.7 Scenario 3: High demand profile

The average holding time per line and controlled stop for scenario 3 is shown in Figure 4.24 as a difference from that of the Benchmark case. Given the small share of synchronisation decisions by the first two controllers and in the absence of tap-in data for the other two, the holding time at the transfer stop is short and yet it introduces a disturbance in the operation of line 17, which necessitates holding to restore regularity at both downstream controlled stops. When tap-ins are available to C3 and C4, all vehicles opt in favour of synchronisation and, given the ignorance of the effect of the disturbance on vehicle arrival times, the average holding time at the transfer stop is almost 4min and the restoration of regularity downstream requires longer holding times.
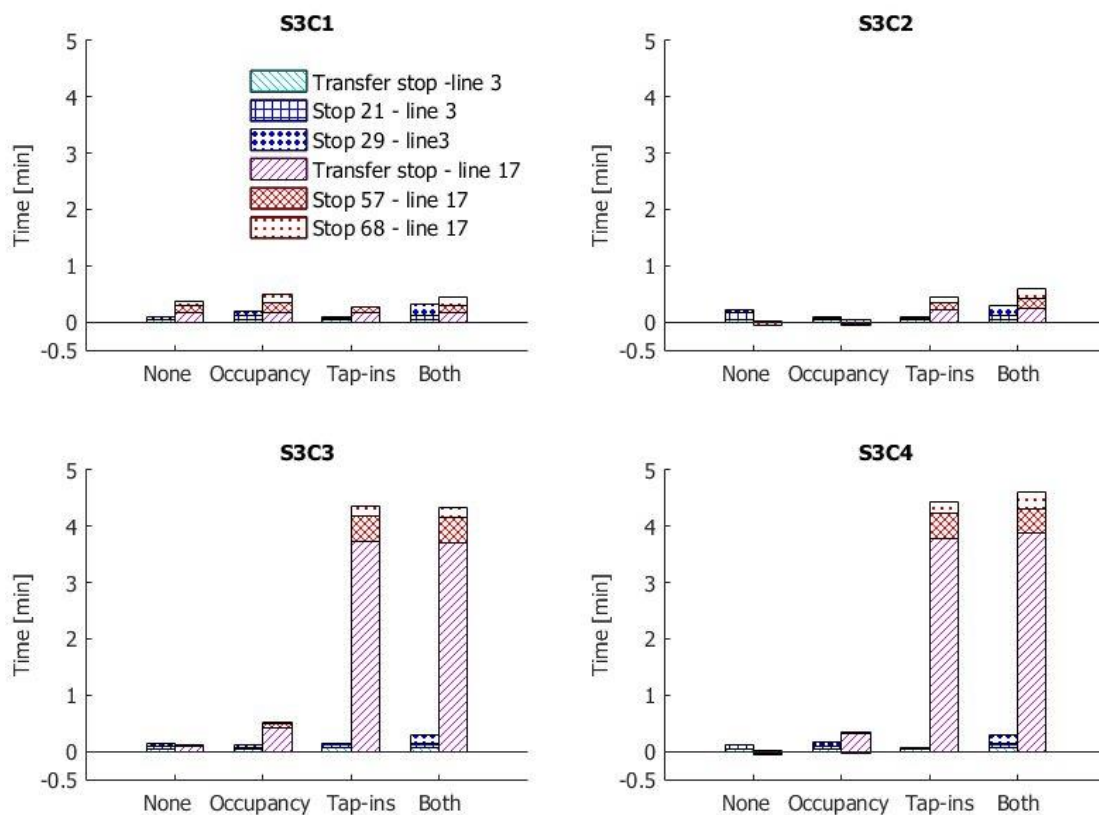

Figure 4.24: Difference compared to the Benchmark case in the average holding time per line and controlled stop for scenario 3.

The results for the perceived trip time from origin to destination per passenger in each passenger group are visualised for the four controllers in Figure 4.25 to Figure 4.28. What is remarkable in this scenario, is the achieved reduction in the waiting time due to denied boarding for downstream passengers and at the transfer stop for those transferring. The former leads to savings of up to 1min per passenger and could be attributed to a vehicle synchronising, receiving the transferring passengers and, thereby, leaving more space to the next one for those being generated downstream. Even though not many transfers are synchronised, the transferring passengers save up to 4min per person in their total perceived trip time and the increase in the holding time for traversing passengers is negligible.

The only deviation from this pattern occurs when tap-ins are available to C3 and C4, in which case all transfers are synchronised. Due to the high volatility in the actual vehicle headways, a synchronisation decision leads to long holding times at the transfer stop and increases the waiting time of downstream passengers. Transferring ones need to wait less for their connection, but experience a longer riding time. Although these controllers are designed to take into account the on-board crowding and prevent an uneven distribution of the demand among vehicles, they fail to do so in this case. The reason behind this result could be either the short horizon that creates a myopic, inadequate view of the network load in real-time, or the quality of the predictions in combination with the availability of real-time data, which are also a function of the horizon.

It is worth noting that in this high demand scenario the performance of controllers C1 and C2 is different. This is reasonable since the capacity constraints are activated and the denied time component in the generalised cost function becomes significant enough to influence the controller decisions.
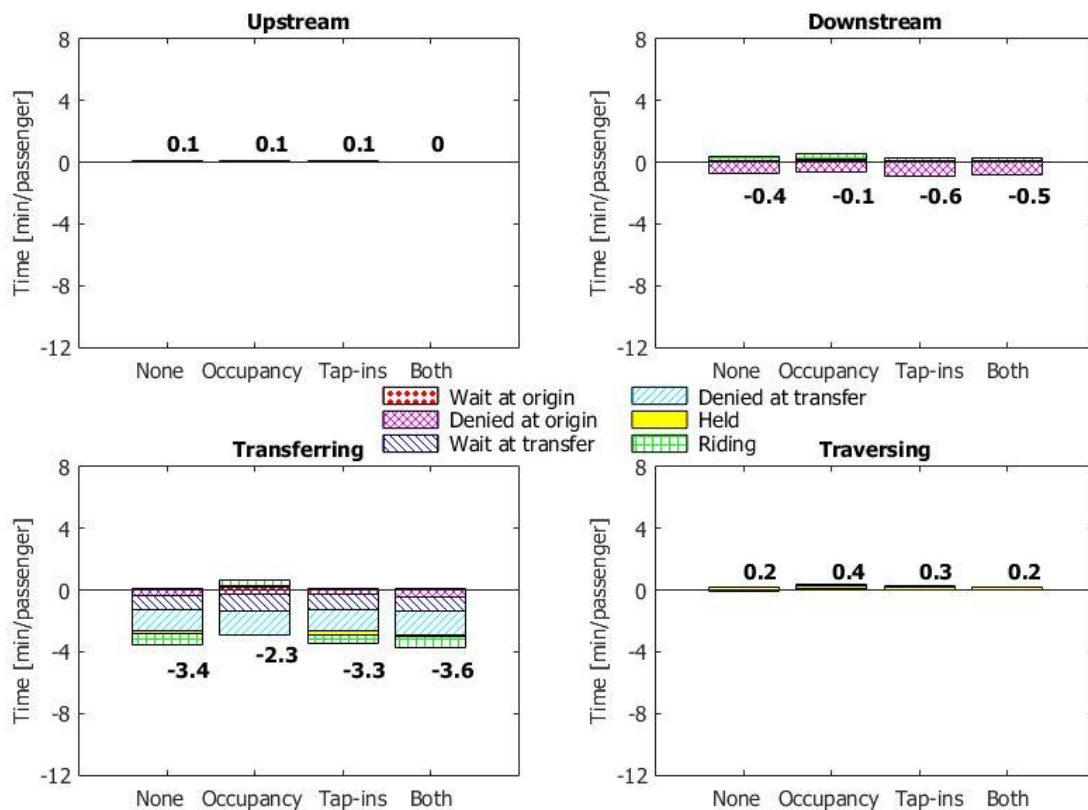

Figure 4.25: Difference compared to the Benchmark case in the perceived trip time from origin to destination per passenger in each passenger group for the case S3C1.

Figure 4.26: Difference compared to the Benchmark case in the perceived trip time from origin to destination per passenger in each passenger group for the case S3C2.

Figure 4.27: Difference compared to the Benchmark case in the perceived trip time from origin to destination per passenger in each passenger group for the case S3C3.
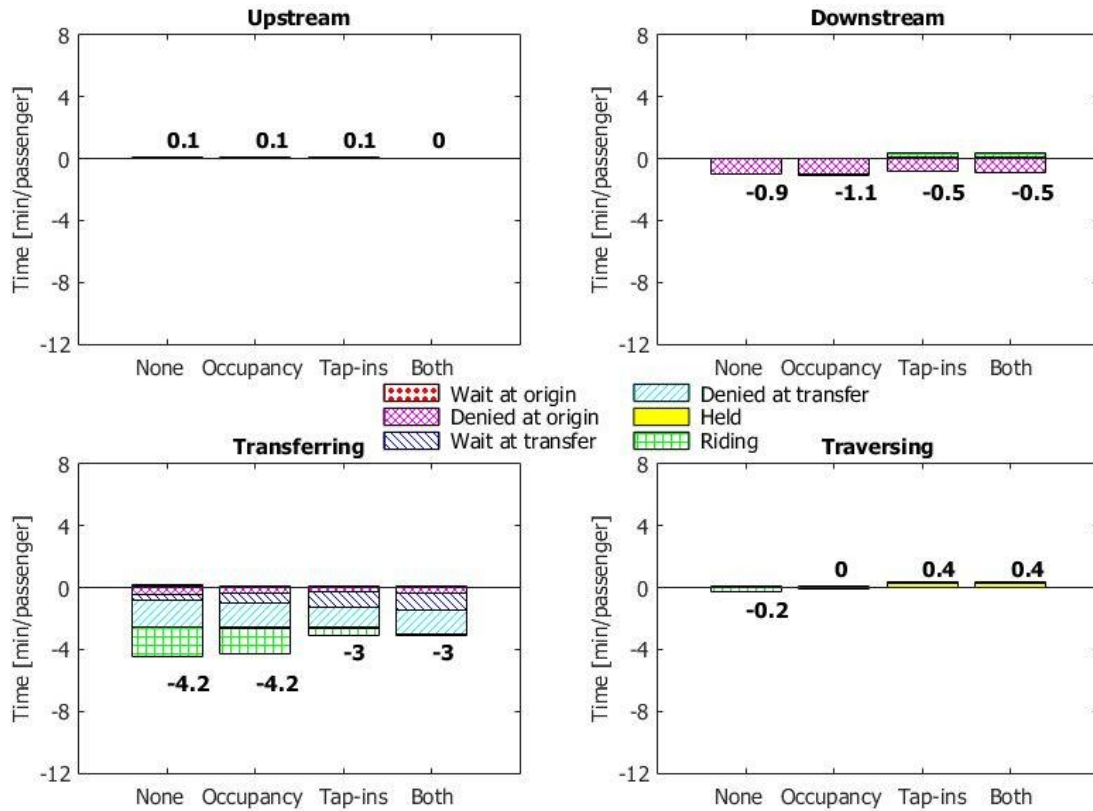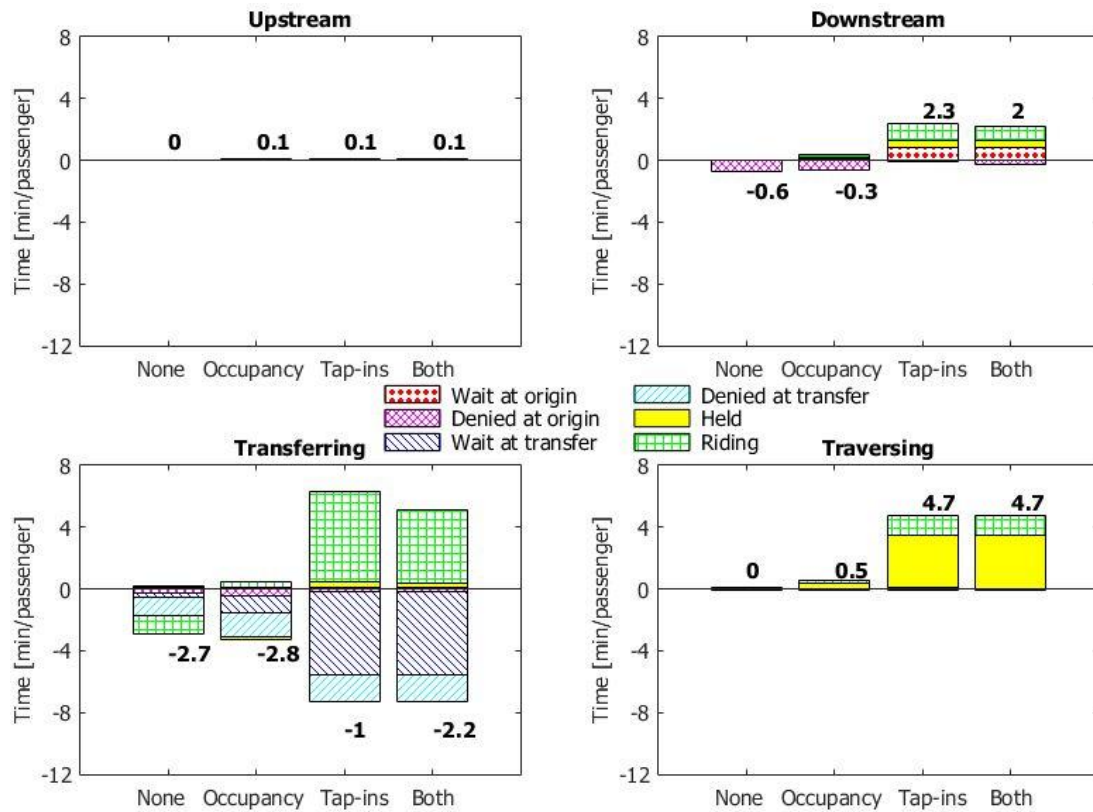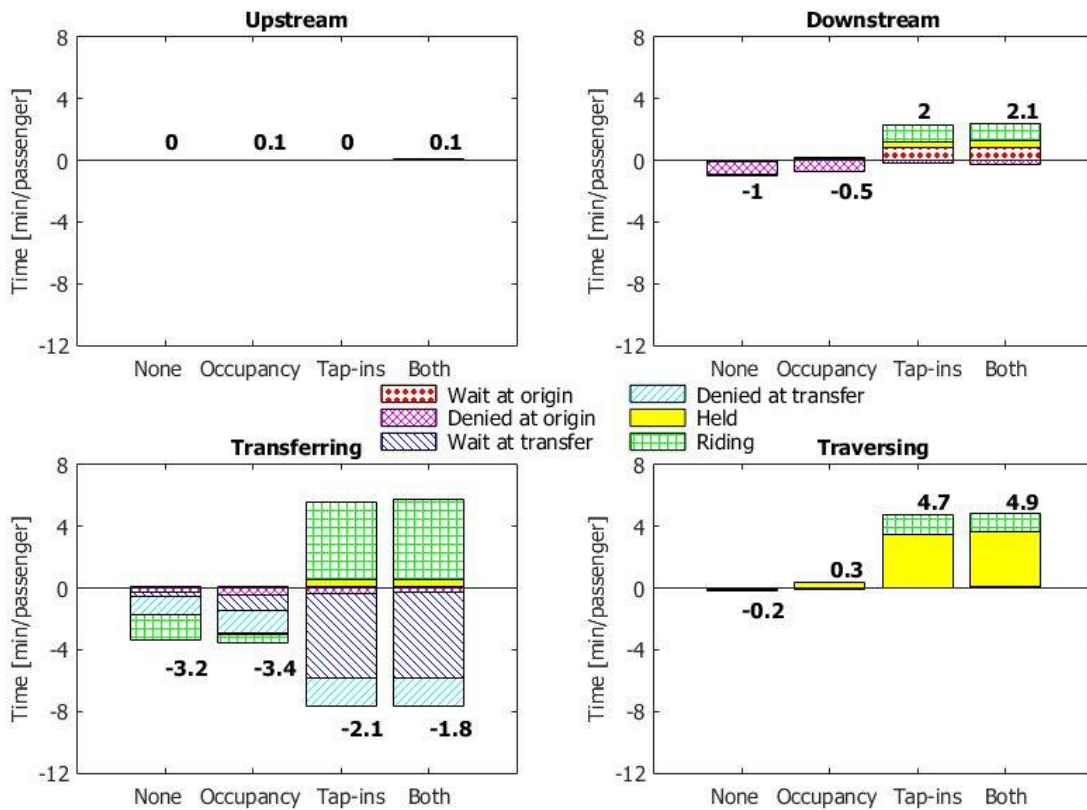
Figure 4.28: Difference compared to the Benchmark case in the perceived trip time from origin to destination per passenger in each passenger group for the case S3C4.

Based on the results on the transfer waiting time distribution for scenario 3, Figure 4.29, it can be seen that, for the Benchmark case, 2% of the transferring passengers are denied boarding. There is even a 1% probability that they will be denied boarding by multiple vehicles, since the maximum transfer waiting time is 45min. The step that was visible in the Benchmark curve of Figure 4.13 which clearly distinguished two transferring passenger groups fades as the demand rises. This stems from the randomness in the vehicle arrivals and the unreliability of the service.

In the controlled cases the maximum waiting time drops to 25min, yet the probability for a direct transfer is only 3%. When tap-ins are available to C3 and C4, this percentage amounts to 22%. This, however, comes to the expense of their riding comfort and increases the perceived trip time of the other passenger groups.

Figure 4.30 demonstrates once more the service unreliability, since the additional holding time for synchronisation can be up to 4.5min when the share of synchronisation decisions exceeds the 10%. As discussed in the service reliability results of subsection 4.3.2, the operation of line 3 becomes very irregular and, therefore, the arrival time predictions are more uncertain and less reliable. The controllers do not take this into account, however, and with knowledge regarding the arrival time and volume of passengers, they judge that the vehicle should be held for synchronisation. Yet the time that it actually needs to wait is much higher, leading up to a regretted decision, which in some cases creates vehicle bunching.
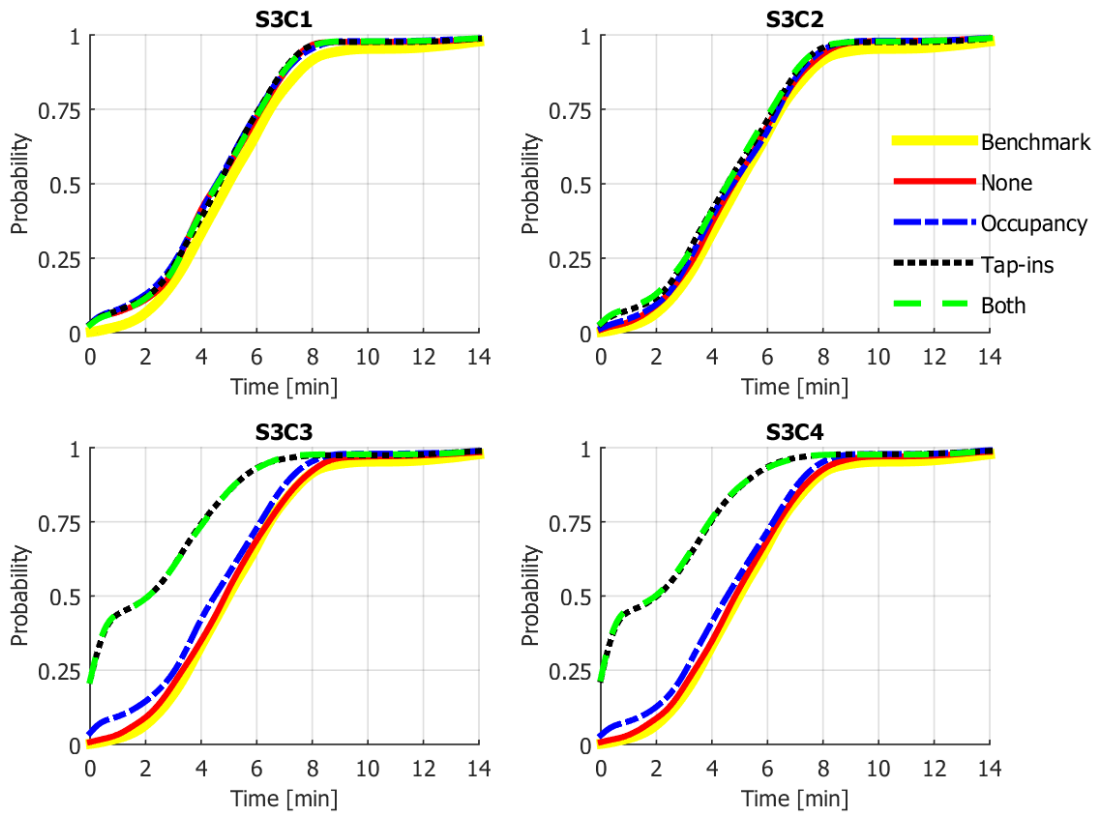
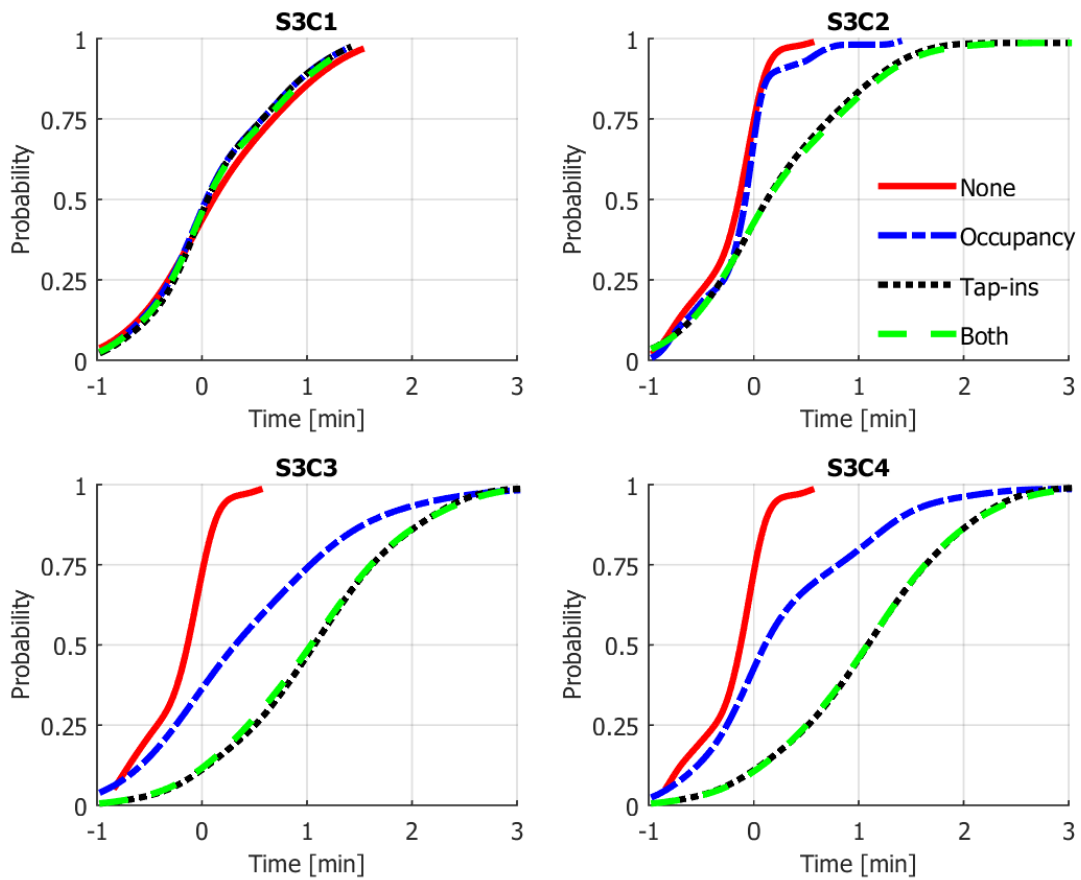Figure 4.29: Cumulative distribution of the transfer waiting time for scenario 3.

Figure 4.30: Cumulative distribution of the unexpected holding time for synchronisation for scenario 3.

The 90th percentile of the total vehicle trip time difference is given in Figure 4.31. In this scenario, the effect on line 3 vehicles is negligible, while the increase observed for line 17 is less than 2min in most cases. When all vehicles are synchronised and the extra average holding time is 5min, the 90th percentile of the trip time of line 17 is 7min longer than that of the Benchmark case.
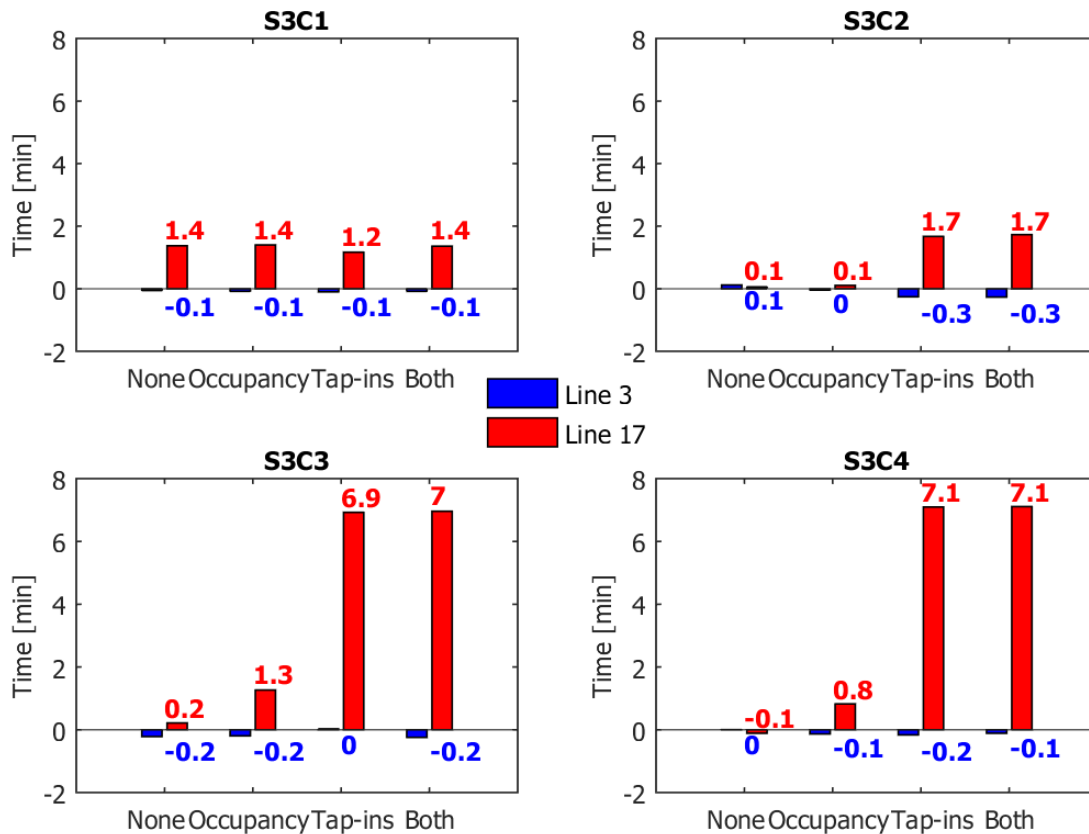
Figure 4.31: Difference in the 90th percentile of the vehicle trip time per line for scenario 3 compared to the Benchmark case.

These results for the third and fourth controller when tap-ins are available are found to be quite unsettling, which calls for further investigation of their causes. These controllers aim to even out the distribution of the transferring passenger demand among vehicles, yet in the presence of tap-ins they end up synchronising all transfers. A primary suspect for this peculiar performance is, as previously mentioned, the choice of the horizon length, which may create a myopic view of the situation in the network.

In order to test this hypothesis, a new scenario is constructed, hereafter to be referred to as S4, which uses the demand profile of scenario 3 but includes all stops in the horizon. This is controlled using the third controller, C3. The reason for this selection is the fact that the introduction of uncertainty in the predictions in C4 did not improve the performance and required longer running times.

Figure 4.32 visualises the results of this case for the first KPI, which are encouraging. It is shown that there is value in having real-time passenger data since the transferring can save 3.7min and less passengers downstream are denied boarding, while the effect on traversing passengers is negligible.
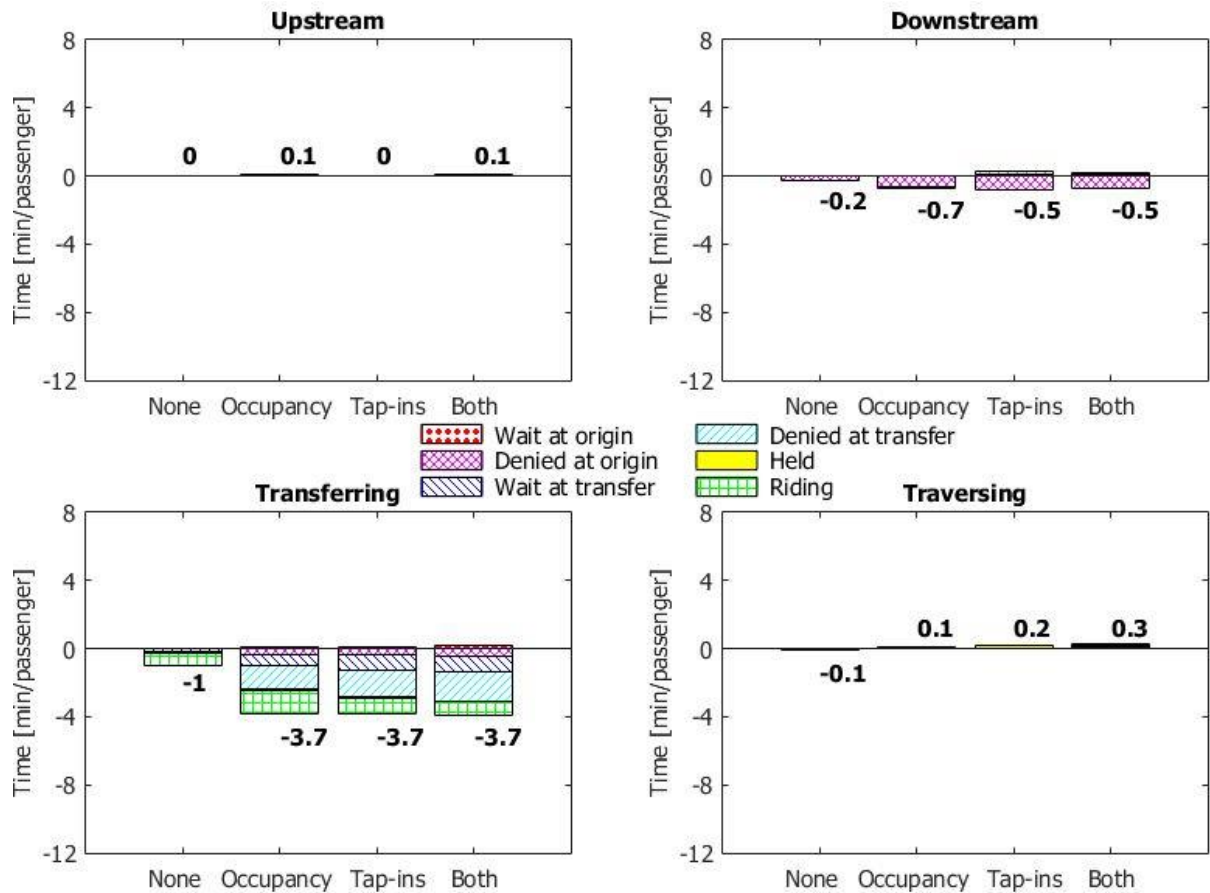
Figure 4.32: Difference compared to the Benchmark case in the perceived trip time from origin to destination per passenger in each passenger group for the case S4C3.

The results for the other KPIs are shown in Figure 4.33. Interestingly, in the case of the long horizon in this high demand scenario controlled by C3, the availability of tap-ins does not result in 50% synchronisation decisions but merely 8%. This proves that the highest share of synchronisation decisions does not lead to the greatest time savings for transferring passengers.

When plotting the control effect on downstream passengers per stop for line 17 and including this fourth scenario with the long horizon (Figure 4.34), the claim made in subsection 4.3.3 that a longer horizon would have yielded better performance, is confirmed.

By comparing the results of different PRTD levels, the conclusion can be drawn that once tap-ins become available in this high demand scenario, the benefits are either compromised or unaffected, which means that the vehicle occupancy is the most valuable real-time passenger data source.

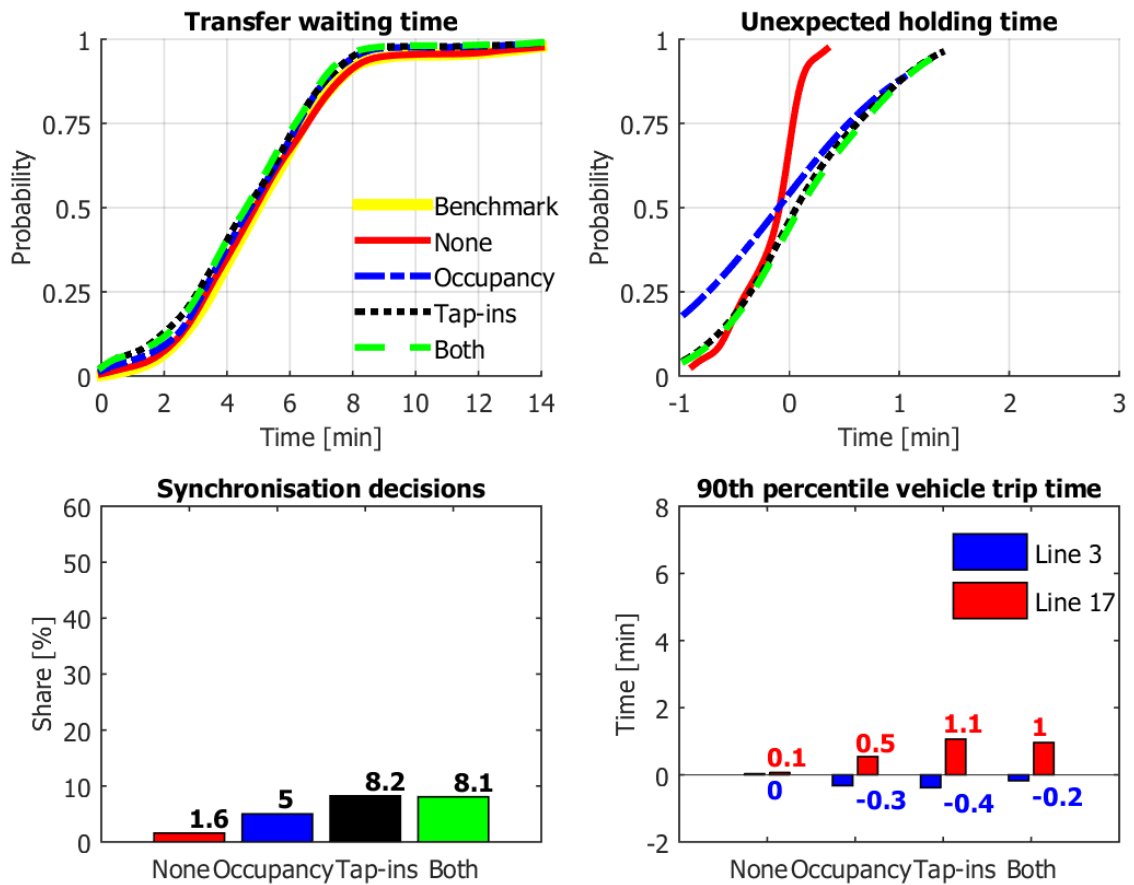*Figure 4.33: Results for the high demand scenario and the C3 controller, while considering a longer horizon (case S4C3).*
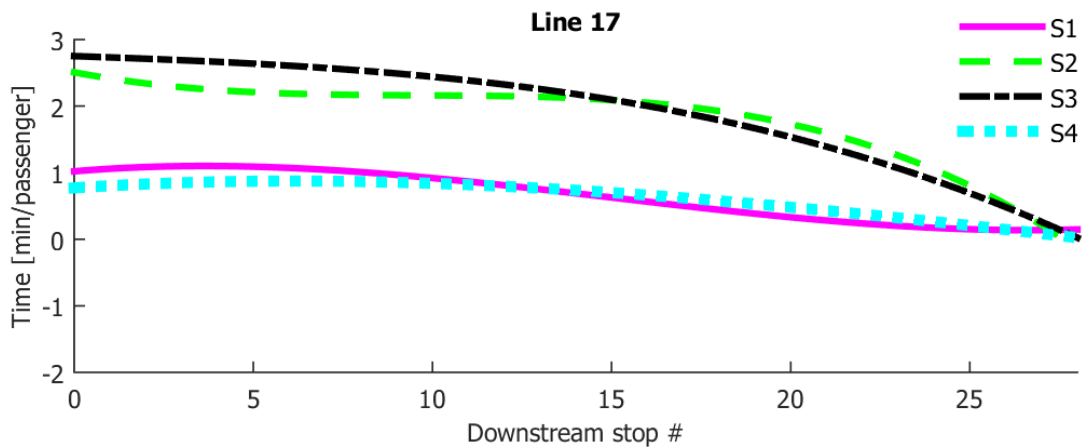

Figure 4.34: Difference from the Benchmark case in the perceived total time for downstream passengers per stop and scenario, averaged over all controllers and passenger data levels for line 17.

5 Conclusions

The present research investigated the benefits of including different types of real-time passenger data in the real-time holding control strategies of public transport vehicles. In doing so, four controllers were developed that could utilise such real-time data. The performance of these controllers was assessed by simulating transit operations with the BusMezzo simulation model for a case study network. Four cases were distinguished regarding the available passenger real-time data types, namely: (1) no real-time passenger data; (2) real-time vehicle occupancy; (3) real-time passenger tap-ins; (4) combination of (2) and (3). The controller variants differed in their underlying assumptions, but they all apply a rule-based strategy for transfer synchronisation that aims at selecting the holding time that yields the lowest total generalised passenger travel times.

The first controller ignored vehicle capacity constraints, which were added in the second. The third one further increased the complexity by taking into account the effect of on-board crowding on the passenger riding experience. The last variant introduced uncertainty in the passenger demand levels by drawing values from a distribution instead of using the historic averages.

Section 5.1 summarises the findings of the case study, section 5.2 answers the research questions and section 5.3 presents the main contributions of this research. These lead to the practical recommendations discussed in section 5.4. Following those, section 5.5 reflects on the choices made in the controller development. The chapter finishes with the proposal for future research in section 5.6.

5.1 Main findings

The case study examined two lines of a real public transport network with a single transfer stop under three demand levels. Each of them was simulated using all combinations of the four controller variants and the four levels of passenger real-time data. The benchmark was a controller which considered only holding control for headway regularity and not for transfer synchronisation.

The performance assessment of the developed controllers in comparison to that of the benchmark proved that there is value in considering to use holding to synchronise transfers. The highest benefit was achieved when the demand level was low, more vehicles could be synchronised and transferring passengers saved up to a full headway (10min) in transfer waiting time, while on-board passengers experienced a delay of 1-2min. However, as the demand rose, the highest share of synchronisation decisions did not result in the most savings for transferring passengers.

In high network loads, the line operation becomes more irregular and the real-time control of vehicles should focus more on re-establishing regularity rather than looking into the

cooperation with other lines. Because of the assumption made by the controllers that the vehicle arrival time predictions are fully reliable, this irregularity has been ignored. The consequence of that is the underestimation of the expected holding time for synchronisation, which led the controllers to regrettable synchronisation decisions and in some cases created vehicle bunching.

With respect to the four controller variants, it has been shown that the first two controllers perform similarly, except for in the case that the demand is high and passengers may be denied boarding. In this case, the controller with the capacity considerations was more conservative in deciding to synchronise which yielded a better performance, i.e. one additional minute saved for each passenger transferring and waiting downstream. The inclusion of on-board crowding influenced the controller performance at all demand levels, since over a certain vehicle load the passengers are more comfortable waiting for the next vehicle rather than standing on-board. The introduction of uncertainty in the passenger flow predictions by the fourth controller did not affect the performance. This could be attributed to the meshing resolution, since the 5% steps of drawn percentiles could be argued to be too coarse. Yet a finer mesh would lead to a further increase of the running time, which is already 100 times higher than that of the other controllers. Apart from the meshing, another argument could be that within each draw from the distribution, the same degree of uncertainty, i.e. the same percentile, is applied to all passenger flow predictions. As a result, the direction to which the predictions shift is the same for all stops and lines, which may end up neutralising the overall effect, given that the decisive control variable is the generalised cost difference among passenger groups.

The comparison of the results for the different passenger data types revealed that the vehicle occupancy is the most valuable real-time passenger data source and it is best exploited when the on-board crowding is considered. Knowing the occupancy of vehicles in real-time can lead to better predictions of the passenger comfort in each control scenario and, hence, better decisions. As the demand rises, the availability of tap-ins leads to more control decisions in favour of synchronisation, which results in a poor controller performance (each downstream passenger needs to wait two more minutes, while traversing passengers are held up to 5min each). This is attributed to the short horizon (10 stops downstream of the transfer stop) adopted by the controllers, which creates a myopic view of the network load. Extending the horizon up to the end of the line expanded the view of the load over the network and resulted in more informed and, therefore, improved control decisions (almost 4min and 1min saved per transferring and downstream passenger, respectively).

5.2 Answers to research questions

In this subsection the formulated research questions are answered. The sub-questions helped to gain supportive knowledge for answering the main question and are, therefore, discussed first.

1. *According to the literature, how are passengers included in the public transport operational control?*

   The review of holding control strategies revealed that passengers were only included in the performance evaluation in early studies, while more recent ones base the control decision on its expected effect on passengers. This effect is quantified by looking at different passenger streams with conflicting interests, such as passengers traversing a control point or being generated downstream of it. The amount of downstream stops considered depends on whether a local or a global perspective is adopted and it is determined by the choice of horizon length. In studies with intersecting lines, where the synchronisation of transfers is decided, the transferring passengers are also distinguished. When vehicle capacity constraints are applied, an additional stream is taken into account, namely that of passengers who were denied boarding.

2. *According to the literature, how can prediction methods make use of real-time passenger data?*

   The literature study showed that this research field is still in its early stages. The predictions are mainly focusing on the vehicles and their arrival times, while some studies attempt to infer the passenger activity based on the vehicle movements. Only few studies make use of real-time passenger data, which in most cases corresponds to the vehicle occupancy. In those cases, the passenger demand forecast relies on the historic averages regarding the passenger arrival, alighting and transferring rates and applies them on the real-time occupancy. When sensors are assumed at stops, tracking the arrival, boarding and alighting of passengers, the studies apply statistical methods, such as ARIMA, to update the historic estimates with the online data.

3. *How can operational control strategies be enriched using real-time passenger data?*

   The prediction method used in this study is based on the historic estimates, which are replaced by the actual values up to the point that real-time data is available. Using this approach, the prediction is only made for a smaller time horizon, thereby decreasing its uncertainty. This scheme is, then, applied in a rule-based controller which selects a holding control strategy to either synchronise transfers or simply regularise the service, depending on the expected impact of each option on passengers.

4. *Which are the most appropriate performance measures for the effect in question?*

Since the objective of the proposed controllers relates to the passenger travelling experience (i.e. minimising their perceived travel time), and especially looks into the synchronisation of transfer, the key performance indicators which are deemed most appropriate concern the passenger activity. These are the following:

a) The perceived trip time from origin to destination per passenger. It is evaluated separately for four passenger groups, since the effect of the control strategy differs among them. The groups use the transfer stop as reference and correspond to passengers (1) alighting upstream, (2) generated downstream, (3) transferring and (4) traversing the transfer stop.

b) The distribution of the transfer waiting time. It checks to what extend the envisaged reduction is achieved.

c) The distribution of the unexpected holding time for synchronisation. This is an indicator of the vehicle arrival time prediction reliability.

d) The 90th percentile of the vehicle trip time per line. It assesses the impact on vehicle operations and is commonly used in determining the fleet size and might be a decisive property for public transport authorities and operators regarding the actual implementation of the strategy.

5. *How does the performance of the developed controller vary under different conditions regarding the passenger demand distribution and service reliability?*

As the demand rises, the service reliability deteriorates and less synchronisation decisions are appropriate. Due to the assumption made in the controller development that the vehicle arrival time predictions are fully reliable, the controllers did not account for this headway volatility and for this reason, fail to prevent regrettable synchronisation decisions. However, if the vehicle occupancy is known in real-time, then the controller has a better estimate of the network load and its decisions lead to time savings for transferring passengers while causing a small delay for those being held at the transfer stop.

Having answered the sub-questions, an answer can also be given to the main research question: '*What is the effect on the performance of an urban transit network, when different types of real-time passenger data are used in its holding control strategy?'*

The effect on the performance caused by the different real-time passenger data types is mostly noticeable when the controller takes into account the on-board crowding. In this case, knowledge of the vehicle occupancy leads to decisions that achieve time savings for transferring passengers and cause only a small delay for those being held at the transfer stop. The availability of tap-ins results in the favouring of synchronisation decisions, which is undesirable in high demand loads and can be contained by adopting a longer horizon. When they are both available, the attainable benefits are either compromised or unaffected, which means that the vehicle occupancy is the most valuable real-time passenger data source.

5.3 Main contributions

In the research field of real-time public transport control a shift has been noted from vehicle-based to passenger-based control strategies. Most of the studies use historic data on passenger activity to reach a control decision, while some are making use of real-time data depending the measuring equipment assumed available. However, public transport systems are increasingly equipped with different types of passenger data that can be transmitted in real-time. Decisions concerning their deployment and availability should be based on scientific results and to the best of the author's knowledge, no such study had been conducted, where the attainable benefit from the different types of data is investigated.

The main scientific contribution of this study is the inclusion of the passenger level of comfort in the generalised cost function of the controller, which achieves the evening out of the transferring passenger demand among the vehicle fleet. Moreover, the controller can make use of different real-time passenger data sources and adjust its passenger flow prediction model based on the data that is available.

In terms of practical contributions, the developed controller has shown the value of considering the synchronisation of transfers for lines whose schedule is designed such that the vehicle arrivals are synchronised. By applying it, the transferring passengers can save up to a full headway of waiting time at the transfers stop, while only a small delay is experienced by passengers held at the transfer stop. Another contribution is that the controller is applicable in real-time, since its running time is fractions of a second. Last but not least, the conclusion that the vehicle occupancy is the most valuable source to acquire in real-time is valuable for public transport authorities and operators.

Given these contributions, it can be stated that the research objectives of this study have been successfully met.

5.4 Practical recommendations

The conclusion has been drawn that the consideration of transfer synchronisation in the real-time holding strategy is beneficial in cases where the scheduled vehicle arrivals of intersecting lines are synchronised. Additionally, the vehicle occupancy has been concluded to be the most valuable source of information to acquire in real-time. Based on these findings, recommendations can be made for the public transport authorities and operators.

The first suggestion concerns the controller selection. It is advised to use a controller that takes into account the riding time component in the generalised passenger cost function. This component is based on the on-board crowding conditions, estimates the comfort level of passengers on-board and makes decisions that improve the level of service and spread the transferring passenger demand evenly across the fleet. Given the increased running time of the fourth controller, when the uncertainty in the passenger flow predictions is included, the third controller variant developed in this study is the one that is proposed to be used in

practice. An important parameter to be set in the selected controller is the horizon length. When the demand level is expected to be high, a long horizon should be chosen in order to prevent a myopic view of the network load.

Regarding the type of data whose collection in real-time should be prioritised by the public transport authorities and operators, the study points towards vehicle occupancy. This is nowadays available through APCs, equipment placed on-board vehicles. However, as argued by Moreira-Matias and Cats [2016], APCs are not yet placed in all vehicles of the fleet and their records are usually not communicated in real-time. This would have to be adjusted to provide measurements in real-time and the operators should choose the lines they wish to coordinate and equip all their vehicles running during the coordination period with APCs.

A practical consideration that may impede the implementation of this holding controller is the situation in which the transfer stop is located upstream of a traffic signal which is designed to unconditionally give priority to approaching public transit vehicles. The conflict arises from the fact that the vehicle sensor detects the presence of one, yet the time it reached the signal is dependent on its holding time at the stop. As suggested by Daganzo and Anderson [2016], a solution would be to apply conditional signal priority. This could be conditioned to be activated once the holding time has elapsed, or, in case of a transfer synchronisation decision, request priority when the next public transit vehicle is detected.

With respect to the location along the route of other control points which only focus on restoring the line regularity, it might be valuable to have at least one upstream of the transfer stop. This is in line with literature findings stating that services with higher uncertainty require more control points [Cats et al., 2014]. The improvement of the line regularity is expected to make the vehicle arrival time predictions more reliable and therefore, lead to a better performance of the controller at the transfer stop.

A point of interest for practical implementation is the passenger composition for which it is worth adopting a transfer synchronisation strategy. Even though such an analysis was beyond the scope of the present study, it can be stated that lines whose schedules are designed to synchronise the vehicle arrivals at a stop are good candidates, since such a design denotes that the amount of transferring passengers is significant. Further research is required before more insightful recommendations can be made on this matter.

Last but not least, it is recommended to treat carefully situations where one line is desired to be coordinated with multiple ones, each having a different transfer stop. The developed controller ignores decisions made at other locations and it should, therefore, be adjusted to take them into account and possibly prioritise among them, according to the importance of securing each of the transfer connections.

5.5 Limitations

During the analysis of the application results, some of the limitations of the controllers have been highlighted and are summarised in this section to give directions for ways to improve them.

The most important limitation has been proven to be the negligence of the uncertainty in the vehicle arrival time predictions. Based on the findings of Daganzo and Anderson [2016], taking this uncertainty into account when the service reliability is low significantly improves the performance and should, therefore, be included.

In some cases, the myopic view of the network conditions has been deemed responsible for the poor performance of the controller. It has been suggested for these instances to make use of an extended horizon, which looks at more downstream stops. However, in highly saturated network conditions, the assumption made in the controller development that passengers who are denied boarding will be served by the next two vehicles of their line, might not hold. This means that the root of the problem might not be the number of stops considered, but rather the amount of vehicles. A countermeasure for this drawback would then be to extend the horizon of vehicles considered by the controller, depending on the expected network load.

As either of the horizons (downstream stops or affected vehicles) is extended, the reliability of predictions diminishes along the dimensions of space and time. Therefore, assigning the same reliability to all of them is a limitation of the current model. A mechanism to resolve this could be the implementation of smaller weight based on the reliability reduction, as demonstrated by Zhao et al. [2003].

In terms of real-time applicability, the inclusion of uncertainty in the passenger flow predictions in the controller resulted in running times in the order of seconds, 100 times longer than the other controllers. This controller should be further developed to make it applicable in real-time. Apart from this, the homogeneity in the implementation of the uncertainty in the passenger flow predictions has been blamed for the neutralisation of the overall effect and the insensitivity of the controller performance to its introduction. Adopting a more heterogeneous sampling method, where the degree of uncertainty differs in each prediction, could determine the validity of this claim along with its effect.

On a different note, it is worth discussing the suitability of vehicle arrival times instead of departure times. Even though this approach is justified by the assumption made that the dwell time of a vehicle at a stop is independent of the passenger demand and consequently of the vehicle headway, it might be severely violated when the service reliability is low. A deterioration of the service reliability influences the vehicle headway, along with the passenger demand and the dwell time at stops. In such cases it might be advisable to switch to the expected departure time of vehicles, since it includes the dwell time.

5.6 Future research

The present study has managed to reach its objective and answer its research questions, while paving the way for future studies in the field of real-time public transport control using real-time passenger data. There are two axes upon which this future research could be navigated. The first is the improvement of the developed controllers by treating each of their limitations and the second is to construct more advanced ones based on them. Since the limitations and their possible treatments have already been discussed in section 5.5, this section outlines promising extensions to the developed controllers.

The first expansion could be to allow the re-evaluation of a synchronisation decision in case that the originally expected holding time is insufficient. Instead of a binary choice among the candidates, a cut-off point could be estimated, defining the boundary for favouring a transfer synchronisation decision.

Moreover, the assumption that all transferring passengers require the same transferring time might be lifted in a more advanced controller. As noted by Delgado et al. [2013], the transfer time is not homogeneous. Instead, several paces are expected, meaning that it is possible that some will make the transfer and others will miss it. In order to account for this effect, the authors in that study considered a trapezoidal distribution of passenger walking time, which influences the ability of passengers to transfer within a time interval. As an alternative, Daganzo and Anderson [2016] proposed the implementation of complementary measures, which could encourage transferring passengers to walk faster, such as directional signage and screens with real-time information.

Additionally, as previously mentioned, further research should be conducted in order to conclude on the passenger composition at a transfer stop that justifies the implementation of the transfer synchronisation strategy.

An interesting direction for future research would be the inclusion of a real-time passenger data source that has not been accounted for in the present study, namely the actual destination of passengers. This could be obtained by a fare collection system that requires the specification of the full itinerary in advance. Alternatively, their planned choice of route could be crowdsourced, necessitating the consideration of adherence to the original plan and representativeness of the sample, which could, nonetheless, prove to be valuable as the complexity of the assumed network increases.

In terms of practical considerations, it is worth investigating ways to coordinate multiple lines at one stop or treat more transfer stops within one line or even combine the two in a network that features a common corridor. The value of the different real-time passenger data sources should then be estimated again in order to validate the generalisability of the conclusions made in the present study.

References

Abkowitz, M. and I. Engelstein. 1984. Methods for Maintaining Transit Service Regularity. Transportation Research Record 961: 1-8.

Abkowitz, M. and M. Lepofsky. 1990. Implementing headway-based reliability control on transit routes. Journal of Transportationn Engineering 116: 49-63.

Ambrosino, G. , S. Gini and L. Mussone. 2014. The operation of its in public transport services: guidelines for public authorities and operators. Transport Research Arena (TRA) 5th Conference: Transport Solutions from Research to Deployment. Paris La Défense CNIT.

Barnett, A. 1974. On Controlling Randomness in Transit Operations. Transportation Science 8: 102-116.

Bartholdi, J.J.III and D.D. Eisenstein. 2012. A self-coördinating bus route to resist bus bunching. Transportation Research Part B: Methodological 46: 481-491.

Carrel, A., R.G. Mishalani, N.H.M. Wilson, J.P. Attanucci and A.B. Rahbee. 2010. Decision factors in service control on high-frequency metro line: importance in service delivery. Transportation Research Record: Journal of the Transportation Research Board: 52-59.

Cats, O. 2011. Dynamic modelling of transit operations and passenger decisions. PhD Thesis in the department of Transport Science Division of Transport and Logistics Centre for Traffic Research, KTH Royal Institute of Technology.

Cats, O. 2014. Regularity-driven bus operation: Principles, implementation and business models. Transport Policy 36: 223-230.

Cats, O., W. Burghout, T. Toledo and H. N. Koutsopoulos. 2010. Mesoscopic Modeling of Bus Public Transportation. Transportation Research Record 2188: 9-18.

Cats, O., A. Larijani, H.N. Koutsopoulos and W. Burghout. 2011. Impacts of holding control strategies on transit performance: Bus simulation model analysis. Transportation Research Record, 51-58.

Cats, O., A.N. Larijani, A. Ólafsdóttir, W. Burghout, I.J. Andréasson and H.N. Koutsopoulos. 2012. Bus-holding control strategies. Transportation Research Record, 100-108.

Cats, O. and G. Loutos. 2016. Evaluating the added-value of online bus arrival prediction schemes. Transportation Research Part A: Policy and Practice 86: 35-55.

Cats, O., F.M. Rufi and H.N. Koutsopoulos. 2014. Optimizing the number and location of time point stops. Public Transport 6: 215-235.

Cats, O., J. West and J. Eliasson. 2016. A dynamic stochastic model for evaluating congestion and crowding effects in transit systems. Transportation Research Part B: Methodological 89: 43-57.

Ceder, A. 2007. Public Transit Planning and Operation: Theory. Modeling and practice. 0xford: Elsevier.

Chandrasekar, P., R. Long Cheu and H.C. Chin. 2002. Simulation evaluation of route-based control of bus operations. Journal of Transportation Engineering 128: 519-527.

Cortés, C.E., D. Sáez, F. Milla, A. Núñez and M. Riquelme. 2010. Hybrid predictive control for real-time optimization of public transport systems' operations based on evolutionary multi-objective optimization. Transportation Research Part C: Emerging Technologies 18: 757-769.

Daganzo, C.F. 2009. A headway-based approach to eliminate bus bunching: Systematic analysis and comparisons. Transportation Research Part B: Methodological 43: 913-921.

Daganzo, C.F. and P. Anderson. 2016. Coordinating Transit Transfers in Real Time.

Daganzo, C.F. and J. Pilachowski. 2011. Reducing bunching with bus-to-bus cooperation. Transportation Research Part B: Methodological 45: 267-277.

Delgado, F., N. Contreras and J.C. Muñoz. 2013. Holding for transfers. Transportation Research Board 92nd Annual Meeting. Washington DC.

Delgado, F., J.C. Muñoz and R. Giesen. 2012. How much can holding and/or limiting boarding improve transit performance? Transportation Research Part B: Methodological 46: 1202-1217.

Delgado, F., J.C. Muñoz, R. Giesen and A. Cipriano. 2009. Real-time control of buses in a transit corridor based on vehicle holding and boarding limits. Transportation Research Record 2090: 59-67.

Dessouky, M., R. Hall, L. Zhang and A. Singh. 2003. Real-time control of buses for schedule coordination at a terminal. Transportation Research Part A: Policy and Practice 37: 145-164.

Eberlein, X.J. 1995. Real-Time Control Strategies in Transit Operations: Models and Analysis in the department of Department of Civil and Environment Engineering, Massachusetts Institute of Technology.

Eberlein, X.J., N.H.M. Wilson and D. Bernstein. 1999. Modeling Real-Time Control Strategies In Public Transit Operations. In Computer-Aided Transit Scheduling: Proceedings, Cambridge, MA, USA, August 1997, ed. Nigel H. M. Wilson, 325-346. Berlin, Heidelberg: Springer Berlin Heidelberg.

Eberlein, X.J., N.H.M. Wilson and D. Bernstein. 2001. The Holding Problem with Real–Time Information Available. Transportation Science 35: 1-18.

Fadaei, M. and O. Cats. 2016. Evaluating the impacts and benefits of public transport design and operational measures. Transport Policy 48: 105-116.

Filippi, F., G. Fusco and U. Nanni. 2013. User Empowerment and Advanced Public Transport Solutions. Procedia - Social and Behavioral Sciences 87: 3-17.

Fonzone, A., J.D. Schmöcker and R. Liu. 2015. A Model of Bus Bunching under Reliability-based Passenger Arrival Patterns. Transportation Research Procedia 7: 276-299.

Fu, L., Q. Liu and P. Calamai. 2003. Real-time optimization model for dynamic scheduling of transit operations. Transportation Research Record 1857: 48-55.

Fu, L. and X. Yang. 2002. Design and implementation of bus-holding control strategies with real-time information. Transportation Research Record 1791: 6-12.

Furth, P.G. and T.H.J. Muller. 2008. Optimality conditions for public transport schedules with timepoint holding. Public Transport 1: 87-102.

Furth, P.G. and T.H.J. Muller. 2006. Part 4: Capacity and quality of service: Service reliability and hidden waiting time: Insights from automatic vehicle location data. Transportation Research Record: Journal of the Transportation Research Board: 79-87.

Gordon, J.B., H.N. Koutsopoulos, N.H. Wilson and J.P. Attanucci. 2013. Automated inference of linked transit journeys in london using fare-transaction and vehicle location data. Transportation Research Record: Journal of the Transportation Research Board 2343: 17-24.

Guevara, C.A. and G.A. Donoso. 2014. Tactical design of high-demand bus transfers. Transport Policy 32: 16-24.

Hadas, Y. and A. Ceder. 2010a. Optimal coordination of public-transit vehicles using operational tactics examined by simulation. Transportation Research Part C: Emerging Technologies 18: 879-895.

Hadas, Y. and A. Ceder. 2010b. Public transit network connectivity: spatial-based performance indicators. Transportation Research Record: Journal of the Transportation Research Board: 1-8.

Hall, R., M. Dessouky and Q. Lu. 2001. Optimal holding times at transfer stations. Computers & industrial engineering 40: 379-397.

Hans, E., N. Chiabaut, L. Leclercq and R.L. Bertini. 2015. Real-time bus route state forecasting using particle filter and mesoscopic modeling. Transportation Research Part C: Emerging Technologies 61: 121-140.

Hickman, M.D. 2001. An Analytic Stochastic Model for the Transit Vehicle Holding Problem. Transportation Science 35: 215-237.

HTM, https://www.htm.nl/ (accessed on Apr. 22).

Hwang, M., J. Kemp, E. Lerner-Lam, N. Neuerburg and P. Okunieff. 2006. Advanced public transportation systems: state of the art update 2006.

Ibarra-Rojas, O.J., F. Delgado, R. Giesen and J.C. Muñoz. 2015. Planning, operation, and control of bus transport systems: A literature review. Transportation Research Part B: Methodological 77: 38-75.

Li, J., T.H.J. Muller, H.J. van Zuylen, X. Chen and W. Wang. 2011. Improving the reliability of transit service: Stochastic modeling of holding strategies. Transportation Research Board 90th Annual Meeting.

Liu, G. and S.C. Wirasinghe. 2001. A simulation model of reliable schedule design for a fixed transit route. Journal of Advanced Transportation 35: 145-174.

Liu, T., A. Ceder, J. Ma and W. Guan. 2014. Synchronizing Public Transport Transfers by Using Intervehicle Communication Scheme: Case Study. Transportation Research Record: Journal of the Transportation Research Board: 78-91.

Liu, Y., J. Bunker and L. Ferreira. 2010. Modelling Urban Public Transit Users' Route Choice Behaviour: A Review and Outlook. Rethinking Sustainable Development: Urban Management, Engineering, and Design: 134-146.

Luethi, M., U.A. Weidmann and A. Nash. 2007. Passenger Arrival Rates at Public Transport Stations. Transportation Research Board 86th Annual Meeting.

Ma, Z., J. Xing, M. Mesbah and L. Ferreira. 2014. Predicting short-term bus passenger demand using a pattern hybrid approach. Transportation Research Part C: Emerging Technologies 39: 148-163.

Manasra, H. 2015. Real time control for transit systems with transfers. MSc Thesis in the department of Civil Engineering - Transportation and Highways Engineering, Technion - Israel Institute of Technology.

McCord, M.R., R.G. Mishalani, P. Goel and B. Strohl. 2010. Iterative proportional fitting procedure to determine bus route passenger origin-destination flows. Transportation Research Record: Journal of the Transportation Research Board 2145: 59-65.

Milkovits, M.N. and N.H.M. Wilson. 2010. Simulating bus service reliability: Another use for automatically collected data. 89th Transportation Research Board Annual Meeting. Washington, DC.

Moreira-Matias, L. and O. Cats. 2016. Towards an AVL-Based Demand Estimation Model. Transportation Research Board 95th Annual Meeting Compendium.

Nesheli, M.M. and A. Ceder. 2015a. Improved reliability of public transportation using real-time transfer synchronization. Transportation Research Part C: Emerging Technologies 60: 525-539.

Nesheli, M.M. and A. Ceder. 2015b. A robust, tactic-based, real-time framework for public-transport transfer synchronization. Transportation Research Part C: Emerging Technologies 60: 105-123.

O'Flaherty, C.A. and D.O. Mangan. 1970. Bus passenger waiting times in central areas. Traffic Engineering and Control 11: 419-421.

Sáez, D., C.E. Cortés, F. Milla, A. Núñez, A. Tirachini and M. Riquelme. 2012. Hybrid predictive control strategy for a public transport system with uncertain demand. Transportmetrica 8: 61-86.

Sánchez-Martínez, G.E., H.N. Koutsopoulos and N.H.M. Wilson. 2016. Real-time holding control for high-frequency transit with dynamics. Transportation Research Part B: Methodological 83: 1-19.

Shen, S. and N.H.M. Wilson. 2001. An Optimal Integrated Real-time Disruption Control Model for Rail Transit Systems. In Computer-Aided Scheduling of Public Transport, eds. Stefan Voß and Joachim R. Daduna, 335-363. Berlin, Heidelberg: Springer Berlin Heidelberg.

Student t-Value Calculator, http://www.danielsoper.com/statcalc/calculator.aspx?id=10 (accessed on 06/06/2016).

Sun, A. and M. Hickman. 2008. The holding problem at multiple holding stations. In Computer-aided systems in public transport, 339-359: Springer.

TCRP. 2003. Transit capacity and quality of service manual (TCQSM) 2nd edition. Transportation Research Board. Washington, DC.

Toledo, T., O. Cats, W. Burghout and H.N. Koutsopoulos. 2010. Mesoscopic simulation for transit operations. Transportation Research Part C: Emerging Technologies 18: 896-908.

Turnquist, Mark A. and Steven W. Blume. 1980. Evaluating potential effectiveness of headway control strategies for transit systems. Transportation Research Record: 25-29.

van Lint, J.W.C. 2004. Reliable travel time prediction for freeways: bridging artificial neural networks and traffic flow theory: TU Delft, Delft University of Technology.

van Oort, N., J.W. Boterman and R. van Nes. 2012. The impact of scheduling on service reliability: trip-time determination and holding points in long-headway services. Public Transport 4: 39-56.

van Oort, N., M. Drost, T. Brands and M. Yap. 2015. Data-driven public transport ridership prediction approach including comfort aspects. CASPT 2015: Conference on Advanced Systems in Public Transport, Rotterdam, The Netherlands, 19-23 July 2015.

Wardman, M. 2004. Public transport values of time. Transport Policy 11: 363-377.

Wardman, M. and G. Whelan. 2011. Twenty Years of Rail Crowding Valuation Studies: Evidence and Lessons from British Experience. Transport Reviews 31: 379-398.

Wei, Y. and M.C. Chen. 2012. Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. Transportation Research Part C: Emerging Technologies 21: 148-162.

Weidmann, U. 1994. Der Fahrgastwechsel im öffentlichen Personenverkehr. PhD Thesis in the department of Transport Planning and Systems, ETH Zürich.

Xuan, Y., J. Argote and C.F. Daganzo. 2011. Dynamic bus holding strategies for schedule reliability: Optimal linear control and performance analysis. Transportation Research Part B: Methodological 45: 1831-1845.

Yu, B., S. Wu, B. Yao, Z. Yang and J. Sun. 2011. Dynamic vehicle dispatching at a transfer station in public transportation system. Journal of Transportation Engineering 138: 191-201.

Yu, B. and Z. Yang. 2009. A dynamic holding strategy in public transit systems with real-time information. Applied Intelligence 31: 69-80.

Zhao, J., S. Bukkapatnam and M.M. Dessouky. 2003. Distributed architecture for real-time coordination of bus holding in transit networks. Intelligent Transportation Systems, IEEE Transactions on 4: 43-51.

Zolfaghari, S., N. Azizi and M.Y. Jaber. 2004. A model for holding strategy in public transit systems with real-time information. International Journal of Transport Management 2: 99-110.