UC Berkeley
UC Berkeley Electronic Theses and Dissertations

**Title**
Conditional Signal Priority: Its Effect on Transit Reliability and Transfers

**Permalink**
https://escholarship.org/uc/item/7104w1jz

**Author**
Anderson, Paul R.

**Publication Date**
2017

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

**Conditional Signal Priority: Its Effect on Transit Reliability and Transfers**

by

Paul R. Anderson

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Engineering - Civil & Environmental Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Carlos F. Daganzo, Chair
Professor Michael Cassidy
Professor Paul Waddell

Fall 2017

**Conditional Signal Priority: Its Effect on Transit Reliability and Transfers**

Copyright 2017
by
Paul R. Anderson

## Abstract

Conditional Signal Priority: Its Effect on Transit Reliability and Transfers

by

Paul R. Anderson

Doctor of Philosophy in Engineering - Civil & Environmental Engineering

University of California, Berkeley

Professor Carlos F. Daganzo, Chair

Travel time variability is a perennial problem for both bus transit operators and passengers. On the operations side, travel time variability is caused by variations in traffic effects and station dwell times. These variations lead to a longer cycle time for transit vehicles which increases operating costs. Passengers experience longer out-of-vehicle waiting times when buses are late or headways are unstable. Lateness and instability are also likely to increase the expectation and variance of transfer times, which may be the largest source of travel time variability for transferring passengers. Making matters worse, other works have found that out-of-vehicle waiting time is the part of transit trips that passengers find most onerous. To address the problem of travel time reliability, this work proposes a system of conditional signal priority (CSP) which aims to keep buses on schedule and help them to recover from delays. The use of CSP opens the door for a complementary strategy of transfer coordination, which can further reduce out-of-vehicle waiting time and make transit more reliable for passengers.

Previous papers have considered conditional signal priority, but only in the context of a case study for a particular corridor. No general evaluation has been done, and the potential for reducing travel time variability has been noted but not quantified. The literature on transfer coordination is more developed, but no previous papers have considered multiple connecting buses, uncertain information, or the use of CSP to recover holding delay. Transfer coordination strategies have also been developed for aviation and passenger rail, but these methods are not well suited to buses. Buses have an ability to act independently in a way that these other modes do not.

The strategy of conditional signal priority is based on the rule that buses can request priority at signals only when they meet a certain condition. In the simplest form of the strategy, this condition is that the bus is late. Given this rule, the main control guide is the bus schedule. On a long homogeneous corridor schedules can be defined by their commercial speed (or, inversely, average pace). The schedule pace is treated as a control parameter, with possible values ranging from the pace with priority at all signals (referred to as transit signal priority (TSP)) to the pace with no signal priority (referred to as no control). The

strategy is distributed, so each bus is an independent decision maker which decides whether or not to request priority as it is approaching an intersection. Several assumptions make it possible to consider one bus and one signal at a time. First, the pairing effect is ignored. Second, conflicting requests are possible but are resolved with simple rules and are therefore modeled as random events. Third, there is assumed to be one station per block. The arrival time within each cycle is therefore assumed to be random because of the variability introduced by the intermediate station (which could be skipped or have a random dwell time) and traffic effects. When a request is made, signals respond with green extensions or early greens when possible to accommodate buses, as in most TSP implementations. Buses are assumed to have a queue jump or reserved bus lane at intersections so that they can depart as soon as the signal turns green. This assumption obviates the need to consider queues of other vehicles. Variations of CSP are proposed which use signals to delay buses, select control actions based on the predicted outcome instead of the current state, and introduce a threshold that modifies the condition for requesting priority. It is argued that the schedule deviation of a bus controlled by CSP can be represented by a Brownian motion with two-valued drift representing the cases when schedule deviation is above or below the condition for requesting priority, respectively. The Brownian motion representation leads to formulas for the expected value and variance of schedule deviation in steady state conditions. Similarly, Brownian motion can predict the expectation and variance of distance to recovery after a delay.

Simulations of a single CSP-controlled bus over a long corridor show that a steady state is reached where the Brownian motion formulas accurately predict performance. It is acknowledged that not all bus routes are long enough to reach steady state or to fully recover from a long delay. Nevertheless, a comparison with no control and TSP shows that CSP has better schedule adherence regardless of route length. Simulations of the variants show that using signals to delay buses would be highly disruptive to signal timing and therefore to other vehicles, that selecting control actions based on predicted outcome could avoid overcorrection, and that nonzero threshold values are useful when schedule pace cannot be optimized to balance the two values of drift.

Next, a strategy of transfer coordination is introduced which is made possible by the communications technology and delay recovery mechanism that CSP-controlled buses possess. The transfer coordination strategy is distributed. Each bus looks for the optimal holding time based on available information when it arrives at a transfer point. The optimal holding time is obtained by constructing a cumulative arrival curve for transferring passengers and comparing it to the holding cost, including downstream effects. A case study constructed from observations at a transfer point in Oakland, California shows that this method could reduce transfer delay (the out-of-vehicle waiting time of transferring, through, and downstream passengers) by 30% or more. Simulations show that the strategy never leaves passengers worse off, even with uncertain information.

This work addresses the problem of travel time variability in bus transit systems and its consequences for passengers, namely an increase in out-of-vehicle waiting time due to late buses and missed transfers. The core of this work is a strategy of conditional signal prior-

ity, which is shown to guide buses to a steady state and to provide a mechanism for delay recovery. The delay recovery mechanism, in particular, allows for complementary strategies. To this end, an optimal control strategy for transfer coordination is developed. This strategy advances the state of practice by considering multiple connecting buses, uncertain information, and downstream effects.

Finally, both control strategies have been described in a general way that can be extended to all bus routes that meet the technological requirements. They are designed to run in the background and take action only when needed. Routes where priority is not available at all signals or where the use of priority is constrained by other control strategies are acceptable provided that the bus schedule is adjusted to reflect these restrictions. Similarly, the transfer coordination method can be applied at all transfer points.

This form of conditional signal priority which considers one bus at a time can be generalized to headway-based service. However, there are likely to be additional benefits from considering forward and backward headway when pairing is of concern and a variant of CSP with this feature will be developed in future work.

To my grandfather, Emory L. Kemp, for sharing his love of civil engineering.

# Contents

**Contents** **ii**

**List of Figures** **iv**

**List of Tables** **v**

**1 Introduction** **1**
1.1 Novelty . . . . . 2
1.2 Technology . . . . . 3
1.3 Outline . . . . . 4

**2 Literature Review** **5**
2.1 Conditional Signal Priority . . . . . 5
2.2 Transfer Coordination . . . . . 6
2.3 Objectives . . . . . 7

**3 Conditional Signal Priority for Transit Networks** **9**
3.1 Strategy for constructing technique . . . . . 9
3.2 Rules, parameters of strategy . . . . . 11
3.3 Expected time savings . . . . . 14
3.4 Variants . . . . . 16

**4 Analysis, optimization, & evaluation** **18**
4.1 Scenarios studied . . . . . 18
4.2 Analysis and verification . . . . . 19
4.3 Comparison . . . . . 25
4.4 Macroscopic view and benefits . . . . . 31

**5 Conditional Signal Priority to improve transfer times** **33**
5.1 How CSP enables transfer coordination . . . . . 33
5.2 Delay Recovery . . . . . 34
5.3 Description of control algorithm . . . . . 37
5.4 Evaluation . . . . . 48

5.5 Macroscopic benefits . . . . . 54

**6 Discussion** **56**
6.1 Coordinating pedestrian crossings . . . . . 57
6.2 Unexpected congestion / other constraints . . . . . 58
6.3 Headway-based service . . . . . 59

**References** **60**

# List of Figures

3.1 Priority rule pattern in a grid network, arrows indicate preferred direction . . . 10
3.2 Layout of modeled corridor . . . 13
3.3 Depiction of advance notice and clear lag . . . 15
3.4 Graphical representation of control actions & time saved . . . 16

4.1 Verification of Brownian motion formulas . . . 24
4.2 Root mean squared schedule deviation for different CSP paces . . . 26
4.3 Root mean squared schedule deviation for different CSP paces . . . 27
4.4 Root mean squared schedule deviation for different CSP paces . . . 28
4.5 Phase disruption of different control variants . . . 29
4.6 Schedule deviation with thresholds for different CSP paces . . . 30
4.7 Standard deviation of schedule deviation with thresholds for different CSP paces 31

5.1 Predicted and actual delay recovery . . . 36
5.2 Expected and actual distance to recovery for different initial delays . . . 37
5.3 Possible outcomes with uncertainty . . . 41
5.4 Comparison of the control with and without uncertainty . . . 45
5.5 Arrival curve construction and delay if holding until $A_c$ . . . 47
5.6 Graphical Explanation of Arrival Curve Method . . . 48
5.7 Histogram of walking times . . . 50
5.8 Arrival curve representations of the four scenarios . . . 51
5.9 Equations (5.17) and (5.18) plotted over a range of parameter values . . . 54

6.1 Complementary pedestrian measure on a single route . . . 57
6.2 Complementary pedestrian measure at a transfer point . . . 58

# List of Tables

5.1 Action plan . . . . . 52
5.2 Out of vehicle delay (min) . . . . . 52

## Acknowledgments

First, thanks to the professors I have worked for: David Levinson, for getting me started in transportation research and for excellent grad school advice which I have followed almost to the letter. Nikolas Geroliminis, who taught me much about narrative shaping and presentation and is an inspiration for his gregariousness and sheer productivity. Carlos Daganzo, the best advisor I could ask for, whose approach to research and advising I'll strive to emulate as I start my own career.

Second, I couldn't have made it to this point without the support and encouragement of my friends. A big thanks to all my Minnesota friends, but especially Emily, Marie, Rachel, Josh, Evan, and Dave. My time in Lausanne would not have been the same without Byung-Geun, Kylie, and Lorenzo. Finally, thanks to all of my friends and colleagues in Berkeley. Special thanks to Juan for making introductions, Lewis and Tim for being my sounding boards, David for all his help in planning for the future, and Abhinav for encouragement and shared adventure.

Last but not least, thanks to my parents, Alison and Ed, for support, encouragement, warehousing of stuff, long roadtrips, and much more.

This research was supported by the California Department of Transportation.

# Chapter 1

# Introduction

Travel time variability is a perennial problem for buses that operate on city streets. For transit agencies, travel time variability is a risk that has to be accounted for in their schedule development and fleet management practices (Seward and Taube, 1977). From the passenger perspective, the effects of travel time variability include late buses, bus bunching, and missed transfers (Holroyd and Scraggs, 1966). All of these effects tend to increase out-of-vehicle waiting time, the component of transit trips that passengers find most onerous (Ben-Akiva and Lerman, 1985; Dube et al., 1991; Dziekan and Kottenhoff, 2007; Watkins et al., 2011).

The impact of travel time variability is both large and widespread. All bus transit agencies try to prevent a delay in one bus task from cascading through the schedule. A typical response is to add some buffer time between bus tasks. However, this response lengthens the cycle time, increasing the operating cost and required fleet size for a particular level of service. Since the buffer required to prevent most delays from spreading is a function of the travel time variability, reducing variability would allow agencies to shorten buffers and cycle times. As for passengers, statistics collected by transit agencies and the Federal Transit Administration show that there are approximately 5 billion unlinked transit trips and 1 billion transfers made every year in the US. All passengers are affected when buses are late or begin to bunch due to travel time variability. Transferring passengers face additional transfer delay when buses are not on schedule.

Travel time variability has two main causes: stops at stations and traffic. Buses typically treat all station stops as optional and stop on request. The possibility of skipping stations is a source of variability in itself. On top of this, the dwell time at any given station depends on the number of passengers boarding and alighting, as well as other special circumstances (fare payment, "kneeling", or using the wheelchair lift) (Daamen et al., 2008; Milkovits, 2008; Niu, 2011; Freyss et al., 2013). Traffic signals are the main source of queuing and delays on city streets, and signal delay impacts buses even when there is a reserved bus lane. Buses may even experience more signal delay than other vehicles when the signals on a street are coordinated for a typical car speed since the bus speed will be lower due to stops at stations and slower acceleration (Duerr, 2000; Skabardonis, 2000). Other traffic effects that can be sources of variability include interactions with turning vehicles, parking maneuvers,

commercial loading/unloading, and bicycles.

Several strategies are used in current practice to address these issues. Bus schedules include slack and turnaround time to smooth out small variations in travel time. Measures like off-board fare collection and level boarding are used (mostly in BRT systems) to shorten station dwell times (Tirachini and Hensher, 2011). Transit signal priority (TSP) is used to reduce signal delay for buses (Evans and Skiles, 1970; Jacobson and Sheffi, 1981), and other traffic effects are addressed by reconfiguring street space (e.g. adding turn lanes, removing parking, making a reserved bus lane).

However, most of these strategies myopically focus on reducing delays to buses and do not directly address the things that are more important to passengers. TSP, for instance, does nothing to reduce travel time variability because it reduces signal delay for early and late buses alike. Slack time is a notable exception, though the amount of slack required to smooth out all travel time variability is impractically large (Newell, 1977). As mentioned above, passengers find out-of-vehicle waiting time to be the most onerous component of their trip. Previous studies have also found that passengers place a high value on travel time reliability (Hickman and Wilson, 1995; de Palma and Picard, 2005; Bhat and Sardesai, 2006; Perk et al., 2008; Watkins et al., 2011). For trips where it is important to arrive on time, passengers may be prepared to accept a higher expected travel time if it is more reliable. This occurs because passengers are planning for a worst-case travel time (e.g. the 95th percentile) to limit the chance of arriving late.

In addition, travel time reliability and transfers are increasing in importance. The travel time reliability of buses in mixed traffic is likely to worsen in coming years as vehicle miles traveled in the US are increasing again after a dip in the Great Recession (Federal Highway Administration, 2016). Travel demand has become much more polycentric in recent years (Gordon et al., 1986; Berry and Kim, 1993; Gordon and Richardson, 1996; Arribas-Bel and Sanz-Gracia, 2014; Sweet et al., 2016), yet transportation infrastructure remains oriented toward city centers. Due to this mismatch between design and use, transfers are essential to serve trips that do not visit the city center.

## 1.1 Novelty

Passenger concerns about travel time reliability and out-of-vehicle waiting time could be addressed with dedicated control strategies. A strategy of conditional signal priority is proposed, where traffic signals give priority to buses only when they request it. Buses send priority requests when a certain condition is satisfied, for instance when they are behind schedule. The effects of this strategy are evaluated in two different contexts. First, in the case of a single route, conditional signal priority is shown to guide buses to a steady state where deviations from schedule are bounded and to help them to recover from major delays. Second, in the case of two connecting routes, the ability of conditional signal priority to recover from delays enables the use of a transfer coordination strategy which is shown to achieve large reductions in transfer delay inclusive of holding cost. Taken together, the proposed strategies

of conditional signal priority and transfer coordination have the potential to benefit bus transit by improving travel time reliability and transit accessibility for passengers and by reducing operating costs for transit agencies. Conditional signal priority also benefits other road users by sending fewer priority requests, resulting in less disruption to normal signal timing.

The primary goal of this research is to develop a general method for conditional signal priority. This method is evaluated in two contexts: on a single bus route and on two connecting bus routes. In the first case, the impact of the proposed control strategy on schedule adherence and the ability to recover from major delays are examined. Schedule adherence and recovery from delays are perennial problems for buses and any progress on these fronts is a welcome addition to the literature. In the second case, the impact on transfers is investigated and complimentary control measures for transfer coordination and pedestrian access to bus stations are developed. These issues are important to transit passengers but are frequently neglected in current practice. Active control measures are proposed which advance the state of research and would have significant benefits for transit passengers.

## 1.2 Technology

Finally, this work contributes to a broader goal of making better use of existing infrastructure. In conventional bus operations, travel times, headways, and timed transfers are determined in schedule development and each bus operates independently, checking schedule adherence at various time points along the route. Real-time information is now collected and distributed to passengers, but is not used to its full potential to inform operations. Many facets of the transit passenger experience, including travel time reliability, transfer waiting time, and access to bus stops, could be improved with control measures based on a strategy of conditional signal priority. Conditional signal priority has the same technological requirements as TSP but the proposed control strategy delivers greater reductions in travel time variability with less disruption to other vehicles. The necessary vehicle-to-infrastructure communications technology already exists where TSP is used and could easily be extended to other actuated signals. Transfer coordination supports the ideal of public transit functioning as a connected network as opposed to a collection of independent routes. The proposed strategy has low information requirements and could be applied to inter-agency as well as intra-agency transfers. The necessary real-time information is already collected with existing bus location technology and distributed to passengers through public APIs and screens at stations. Improving inter-agency transfers could be a significant improvement, as many cities have more than one transit operator and transfers between different operators are likely not planned. (As evidence of this, AC Transit system and route maps do not acknowledge bus routes by other operators.) Doing more with existing resources is a worthy goal of all organizations. It is of particular importance for transit agencies because funds for operations are scarce.

## 1.3 Outline

This work is organized into six chapters. Chapter 2 describes previous work and the contribution to the literature. Chapter 3 presents the method of conditional signal priority for transit networks. Chapter 4 describes the analysis, optimization, and evaluation of CSP. Chapter 5 discusses the use of CSP to improve transfer times. Chapter 6 contains notes for implementation, directions for future work, and concluding remarks.

# Chapter 2

# Literature Review

This literature review covers previous work on transit signal priority and transfer coordination, as well as strategies that are used in other scheduled transportation modes to recover from delays and coordinate transfers. In brief, two previous papers have considered conditional signal priority, but only in the context of a case study for a particular corridor. No general evaluation has been done, and the potential for reducing travel time variability has been noted but not quantified. The literature on transfer coordination is more developed, but no previous papers have considered multiple connecting buses, uncertain information, and the use of CSP to recover holding delay. Transfer coordination strategies have also been developed for aviation and passenger rail but are not well suited to buses. Buses have an ability to act independently in a way that these other modes do not.

## 2.1 Conditional Signal Priority

The idea of conditional signal priority was introduced by Janos and Furth (2002), who developed it for a 6.7km (4.2mi) bus route in San Juan, Puerto Rico. Qualitatively, Janos and Furth (2002) made several contributions: they noted that not all TSP activations are useful, developed a simple rule for granting priority, and discussed how the bus schedule should change when CSP is used. However, the control rule was set somewhat coarsely (tolerance from schedule of 1 min) and without an attempt to optimize it. Furthermore, it was only evaluated for the site in question and no general insights or results were offered. Ma et al. (2010) proposed a conditional signal priority strategy with both priority (green extension/early green) and delay (red extension/early red) actions. However, their method assumes knowledge of future vehicle trajectories through the next several signals that is simply not available in practice. Chow and Li (2017) used the same set of priority and delay actions as Ma et al. (2010) to implement a forward headway control similar to Daganzo (2009). This thesis builds on the qualitative arguments of Janos and Furth (2002) and strives to fill a gap in the literature by obtaining general insights and results for a family of conditional signal priority strategies using information that is available. The delay requests

proposed by Ma et al. (2010) and Chow and Li (2017) will be tested as well but we believe that using signals to hold buses is too disruptive to other vehicles to work well in practice.

## 2.2 Transfer Coordination

Transfer coordination plays a smaller role than conditional signal priority in this thesis, but transfers exist in all scheduled transportation while the notion of signal priority exists only for modes which operate on city streets. This literature review covers works about transfers in buses, aviation, and passenger rail.

Most of the existing work on buses uses static strategies that favor transfers. Several works have proposed ways of designing bus schedules to reduce expected transfer waiting times (Hall, 1985; Abkowitz et al., 1987; Lee and Schonfeld, 1991; Bookbinder and Desilets, 1992; Knoppers and Muller, 1995; Hall et al., 2001; Hadas and Ceder, 2010). A few recent papers incorporate real-time information and develop adaptive strategies to help with transfers. A first group use real-time information to take control actions in order to preserve scheduled transfers. Hadas and Ceder (2010) propose using informal strategies to help buses reach transfer points on time. Dessouky et al. (2003) develop holding strategies to preserve timed transfers, but do not examine the effect of uncertainty or consider walking time. Lo and Chang (2012) propose a real time fuzzy bus holding system. A second group look for transfer opportunities in real time rather than focusing on timed transfers in the schedule. Delgado et al. (2013) consider transfer coordination on a heavy rail system using a rolling horizon approach. An innovation of Delgado et al. (2013) is that it considers a trapezoidal distribution of passenger walking time, so it is possible for some to make the transfer while others miss it. Most recently, Nesheli and Ceder (2014, 2015) propose a combination of holding and segment skipping (i.e. skipping several stops in a row) to effectuate transfer coordination.

A similar problem of managing passenger connections exists in aviation, though it is somewhat different because passenger itineraries are known by the airline and airlines are responsible for rebooking passengers who miss their connections. Static strategies of designing schedules to favor transfers, in part by reducing delay propagation, are also used (Hane et al., 1995; Lohatepanont and Barnhart, 2004; Lan et al., 2006; AhmadBeygi et al., 2008; Jiang and Barnhart, 2009; Aloulou et al., 2013; Cadarso and Marín, 2013). In current practice, airlines use adaptive strategies, too. They focus on rerouting aircraft, crew, and passengers in that order after a disruption (i.e. a major delay) (Filar et al., 2001; Bratu and Barnhart, 2006; Arıkan et al., 2016). In academia, adaptive strategies are considered by Bratu and Barnhart (2006), Petersen et al. (2012), and Arıkan et al. (2016). Bratu and Barnhart (2006) optimize operating and passenger delay costs after a disruption using holding alongside flight cancellations, reserve crews, aircraft swaps, and spare aircraft. Petersen et al. (2012) solve aircraft, crew, and passenger recovery after a disruption simultaneously. Arıkan et al. (2016) focus on aircraft and passenger recovery (excluding crew) and include the possibility of increasing cruise speed to make up delays. In general, these strategies

are not translatable to bus operations because transfer coordination and the welfare of passengers are not explicitly considered. Flights may be held for connecting passengers as an indirect means of reducing passenger delay costs, but the methods are focused on constraint satisfaction for aircraft and crew.

The problem of managing transfers also exists in passenger rail. Ticketing policies vary, some rail systems are like aviation and passengers are booked onto specific trains while others function more like public transit where passengers are free to choose their own route. A detailed literature review is given in Cacchiani et al. (2014). Schedule design strategies are also used, e.g. Schaafsma and Bartholomeus (2007); DAriano et al. (2008). As in aviation, the use of active strategies for transfer coordination is highly constrained by infrastructure capacity, rolling stock assignments, and crew schedules (Visentini et al., 2014). Most of the papers in the rail literature are focused on recovering from delays by computing a new disrupted timetable for the entire network in real time that satisfies all constraints (e.g. Chiu et al. (2002), Kanai et al. (2011), Dollevoet et al. (2014)). Schöbel (2001) take this approach but have a step where they decide whether or not timed transfers can be maintained. Distributed approaches are sometimes used, but mainly to manage complexity and a central controller is maintained to ensure constraint satisfaction (Corman et al., 2011). The approaches used in passenger rail have some relevance to buses, but buses are much less constrained by infrastructure capacity and able to act independently, which makes fully distributed control strategies possible.

The transfer coordination strategy developed in this paper differs from those in the literature for buses in that it is less ambitious, requires few data that are readily available, and is simple to apply and evaluate. The method is less ambitious than possible because it does not attempt to control the bus system and instead focuses on one bus and one transfer at a time. The proposed control method can also capture important real-world phenomena that are not considered in other works. These include multiple connecting buses, uncertain information, and downstream effects.

## 2.3 Objectives

As mentioned in the previous chapter, the goal of this thesis is to develop a general method for conditional signal priority. This technological requirements of bus location, actuated traffic signals, and vehicle to infrastructure communications already exist. Simple rules are used for maximum applicability. It creates an opportunity for a complementary method of transfer coordination, which is also presented in general form and with minimal technical requirements for maximum applicability. Finally, both methods improve travel time reliability and out-of-vehicle waiting time, which are the performance measures that matter most to passengers.

The previous works on conditional signal priority were developed for specific corridors. Neither proposed a general method for CSP nor offered insights about the performance that could be expected in other cases. Janos and Furth (2002) satisfies the objective of using

existing technology and simple rules but is not general. Ma et al. (2010) assumes information that is unavailable and Chow and Li (2017) is far too disruptive, so both are impractical to implement. Janos and Furth (2002) reported a small improvement in speed with CSP and no change in travel time variability. Ma et al. (2010) reported significant improvements to bus delay and travel time variability with their strategy, but a large increase in delay to cars. Chow and Li (2017) and Chow et al. (2017) report improvement in bus headway variability and traffic delays but attribute the latter to poor signal timing in the no control case. This thesis aims to show that more progress on these objectives is possible.

Regarding transfer coordination, the previous works that look for transfers opportunistically (Delgado et al., 2013; Nesheli and Ceder, 2014, 2015) meet the goal of wide applicability using existing technology. Some of the papers in the literature use simple rules. Preserving scheduled transfers, as in Dessouky et al. (2003) and Hadas and Ceder (2010), is simple and scalable. Other methods, like the rolling horizon in Delgado et al. (2013), could suffer from scaling problems in large networks. Many papers have struggled with the problem of recovering delays after holding for a transfer. Dessouky et al. (2003) simply prioritizes timed transfers over schedule adherence. Others use informal measures to make up delays (Hadas and Ceder, 2010; Nesheli and Ceder, 2014, 2015). The problem with informal measures like boarding limits and stop skipping is that they benefit on-board and transferring passengers at the expense of other passengers at the stations just before or after a transfer point. Skipped passengers at nearby stations experience an extra headway of out-of-vehicle waiting time. In the long run the travel time variability of trips originating from stations near transfer points would worsen because the impacts of the informal measures are geographically concentrated. The proposed method of transfer coordination is innovative in using signal priority to make up delays. It also has negative effects on downstream passengers at stations near transfer points, but they are much less severe. Downstream passengers experience additional out-of-vehicle time equal to the schedule deviation, which will be constrained to be less than or equal to the backward headway.

# Chapter 3

# Conditional Signal Priority for Transit Networks

A real-time control strategy for conditional signal priority has been developed and is described in this section. The proposed strategy for conditional signal priority uses simple rules to decide when buses request priority, how the signal responds to a priority request, and how conflicting requests are resolved. The purpose of using simple rules at each step is to create a general approach that can be implemented in a decentralized fashion. The rules are general and should apply to inhomogeneous routes and on any network. The advantage of decentralized control is that it can be applied even on very complex networks where centralized strategies would suffer from scaling problems.

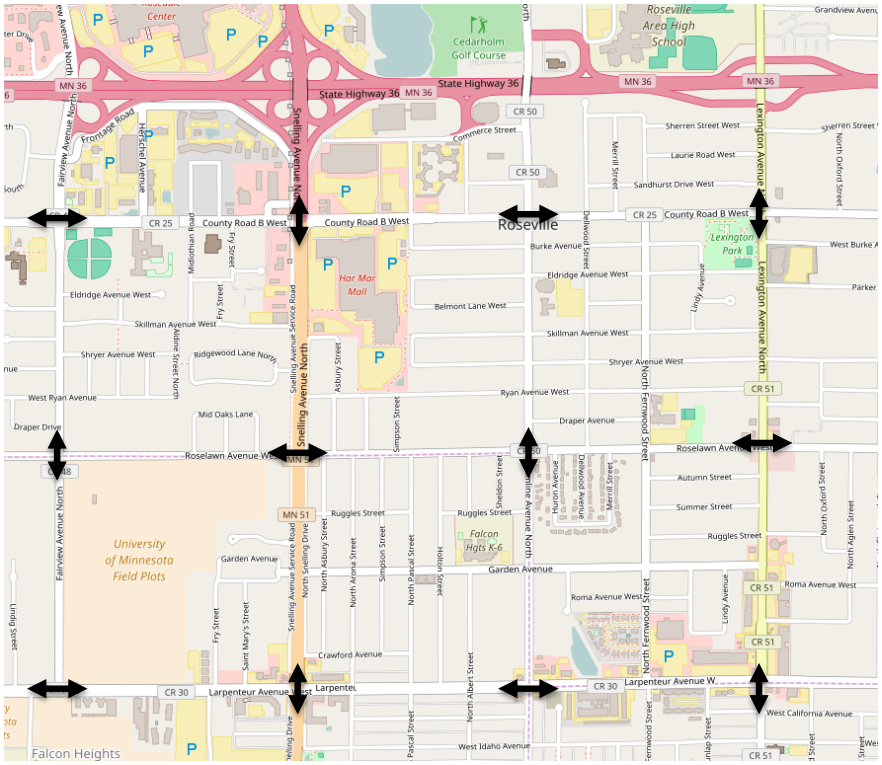
This chapter is organized as follows. Section 3.1 explains the strategy for constructing the technique in broad terms. Section 3.2 describes the rules and parameters of the strategy in detail and explains what buses and signals do. Section 3.3 describes parameters related to the signal timing and shows how these can be used to derive the expected benefit of a priority request. Section 3.4 describes three variants to the base CSP strategy described in sections 3.1-3.3 that will also be evaluated in Chapter 4.

## 3.1 Strategy for constructing technique

In order for the control strategy to be completely decentralized, each bus must be able to make its decision about when to request priority independent of buses on other routes and on the same route. This section explains the rules and assumptions that allow for decentralized control.

When more than one bus route is using signal priority[1], it is possible that a signal will receive conflicting requests. Not all of the requests can be satisfied, so some method of conflict resolution is required. This topic has been the subject of several papers (He et al., 2011; Zlatkovic et al., 2012; Ma et al., 2013; He et al., 2014) which use increasingly complex

[1] "Signal priority" is used in a general sense, as the problem exists for both CSP and TSP.



Source: OpenStreetMap
Location: Roseville, Minnesota

Figure 3.1: Priority rule pattern in a grid network, arrows indicate preferred direction

optimization methods to determine which request should be served. Making an optimal decision is difficult because many factors could be considered (e.g. deviation from schedule, passengers onboard, passengers waiting downstream, upcoming transfers, etc.), some of them are uncertain because they concern future events, and the decision has to be made quickly. A simpler and more predictable way of conflict resolution is to use rules. A first-come-first-served policy is easy to implement and to analyze. Priority is available unless a conflicting request has been made in the same cycle. The probability of conflict can then be estimated from the conflicting route's headway, the cycle length, and the activation rate of the control strategy.

In a grid network, preference could be given to one street over the other in an alternating pattern, as shown in Figure 3.1. Requests from buses on the preferred street would be handled first, while buses on the other street could receive priority if no requests were received from the preferred street. The advantage of this method is that, from the perspective of a single bus, it cuts the number of intersections with potential priority conflicts in half and avoids bias from the schedule when crossing routes have a similar headway.

In aviation and passenger rail, central controllers are used to prevent collisions and man-

age traffic flow in busy areas. Buses, on the other hand, can operate independently because the vehicle properties allow drivers to handle these tasks on their own. In infrequent service, buses are effectively independent of other buses on the same route because passengers tend to time their arrivals to the schedule, hence bunching is not a major issue. When service is more frequent, passengers arrivals are more dispersed and bunching becomes an issue. In the latter case, buses are affected by the position of the bus in front and have an impact on the bus behind. Even so, this dependence can be taken into account using bus-to-bus communication and does not necessarily require a central controller.

For routes with infrequent service, it is entirely appropriate to consider one bus and one route at a time. Priority conflicts with buses on other routes can be modeled as random events. Interactions with other buses on the same route are insignificant and can be neglected. For routes with frequent service, it is possible to implement conditional signal priority without considering the forward and backward headways because the strategy will still react in the correct way to schedule deviations. However, any simulation would need to include multiple buses to correctly model the bus motion.

## 3.2 Rules, parameters of strategy

The premise of conditional signal priority is that a bus is only allowed to request priority if it meets a certain condition. In the most basic form of the strategy, the condition is that the bus is late. This choice of condition means that the main control guide is the bus schedule. The speed of the schedule will be treated as a parameter. The challenge is to select a speed that is high but can be maintained reliably. The range of schedules considered is from no control (the slowest) to priority at all signals (the fastest, referred to as transit signal priority or TSP). An optional secondary control parameter is a tolerance. Transit agencies, in their performance measurement, consider buses within a certain tolerance of schedule to be on time and a similar policy will be used in the control strategy.

The schedule is defined by an initial dispatch time $t_0$ [s] and a commercial speed $v$ [mph], so the schedule at signal $i$ is $t_i = t_0 + \frac{3600is}{v}$, where $s$ is the signal spacing [mi]. For the purposes of this thesis, $s$ will be constant so that the bus route follows a homogeneous corridor. The layout of a block is shown in Figure 3.2. The schedule can also be defined by its average pace $T_s = \frac{3600s}{v}$. The pace of no control and TSP can be defined as $T_c$ and $T_u$, respectively.

The proposed control strategy is designed to use information that is available and technology which already exists. The only information that the bus needs is the current time and its schedule. To analyze the strategy, additional parameters describing the network geometry, signal timing, and crossing route headway are also required. If the bus decides to send a request, the request is communicated to the signal as a pulse. The pulse does not contain any additional information, but requests can be sent from a predetermined upstream location with predictable travel time (e.g. the nearest upstream station as shown in Figure 3.2 if it is in the same block, or else the upstream end of the block) so that the signal knows when

to expect the bus. Signals must be equipped to receive priority requests and they respond in a predetermined way. In addition, there is assumed to be a queue jump or reserved bus lane so that buses can depart as soon as the signal turns green. Without this assumption, buses would be delayed by queue dispersal at the start of an early green phase.

Similarly, it is useful to have simple rules for resolving priority conflicts because signals have little computational power and only local information. It is true that the technology exists for connected infrastructure to perform online optimization to resolve priority conflicts. However, the complexity and expense that this would entail is unnecessary if the desired outcomes can be obtained with existing infrastructure, local information, and simple rules.

We shall evaluate the performance of the proposed control strategy when it is implemented along the entire length of several intersecting bus routes. The model considers one bus and one signal at a time. This reduction is made possible by several simplifying assumptions:

1. The pairing effect is ignored. This assumption is reasonable for routes with long headways and allows the buses on a route to be analyzed independently.

2. A relationship between bus route headway, cycle length, and the probability of a conflicting priority request is assumed. This assumption allows priority conflicts to be modeled as random events so that the location of buses on other routes is not needed.

3. There is assumed to be one station per block.

4. The variation in dwell times (and other random effects) is large enough that the bus arrives at a random point in the cycle of each signal

## What the buses do

Each instance of the control strategy begins when the bus is departing the station upstream of a signal $i$. At this point, the bus compares the current time, $t_a$, to the scheduled time, $t_i$. If the bus is behind schedule, it sends a priority request to signal $i$. Otherwise, no action is taken. Knowing $t_a$, this control rule can be written:

$$\begin{aligned} &\text{if } t_a - t_i > 0: \\ &\qquad \text{send priority request} \\ &\text{else:} \\ &\qquad \text{no action} \end{aligned} \tag{3.1}$$

The priority request is delayed until the bus is $r$ time units away from reaching the signal, where $r$ is an advance notice parameter that cannot exceed the time to travel the block. For either control action, the model computes the line haul time and the signal delay (with or without priority) and adds these to the current time to determine the arrival time at the

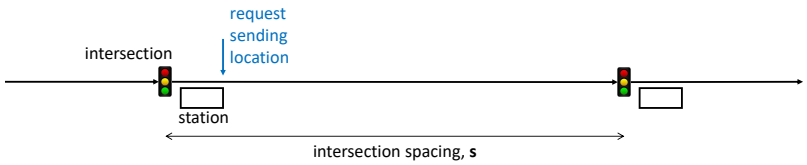



Figure 3.2: Layout of modeled corridor

next station. The variable component of travel time (representing both dwell time at the next station and traffic interactions) is drawn from the specified distribution, added to the arrival time, and the resulting departure time is fed to the next instance, $i + 1$.

## What the signals do

When a priority request is received, the signal first checks for conflicts. Conflicting priority requests are managed using a set of priority rules that define the order in which requests are treated at each intersection. In this thesis, the street network is a regular grid with a signal spacing of $s$ and a bus route spacing of $b$. Consequently, there is a crossing bus route at every signal when $b = s$, the case shown in Figure 3.1. Along any given street, the priority rule alternates between giving precedence to east-west and north-south routes. As a result, any given bus route will have precedence at half of all intersections of bus routes. There remains a possibility of conflict at every other signal when $b = s$, and these conflicts are modeled as random events. The combined headway of crossing routes at intersection $i$, $H_i$, is a part of the bus schedule and is assumed to be known. The probability of encountering a crossing bus is assumed to follow a uniform distribution, meaning that the probability of encountering a crossing bus in the same cycle is $\frac{C}{H_i}$. Not every crossing bus will request priority, but this expression can conservatively be called the probability of a priority request being denied. The true probability would be obtained by multiplying $\frac{C}{H_i}$ by the priority activation rate of a specific control scheme.

More generally, crossing bus routes are encountered at every $\frac{b}{s}$ signals and with an alternating priority rule there is a possibility of conflict at every $\frac{2b}{s}$ signals. In a sparse bus network ($b >> s$) with infrequent service and a priority activation rate less than 1, priority would be available nearly all the time.

## 3.3 Expected time savings

We now evaluate the expected time savings from a priority request. If a priority request is received and there are no conflicts, the signal responds like it would in TSP, attempting to minimize the delay to the bus. A simple two phase timing is assumed for this explanation. There are two control actions that can be taken to speed up buses: green extension and early green. (No action is taken if the signal would already be green when the bus arrives.) To minimize the disruptions to the cycle and to other vehicles, the red phase is divided in half. If the bus is expected to arrive in the first half of the red phase, the signal can perform a green extension. If the arrival is in the second half of the red phase, the signal can give an early green. In the case of a green extension, the next red phase will be of normal length, and the following phase will be shortened to compensate for the green extension. In the case of an early green, the green phase will be of normal length and the following red phase will be extended to compensate for the shortened phase.

The feasibility of both control actions depends on when the request is received. The elapsed time between the time the request is sent and the time the bus arrives, $t_a$, is called the advance notice, $r$. A green extension cannot be granted after the green phase has ended, i.e. if the signal is already red at $t_a - r$. The clear lag, $Y_0$, is the elapsed time between when the signal receives a request for an early green and when the early green begins. It represents the time needed to safely clear the intersection of vehicles and pedestrians. It will be assumed that the clear lag exceeds the advance notice, $r < Y_0$, as this case is more challenging for control, though an early green may be granted late (i.e. after the bus has arrived and is waiting at the signal). Figure 3.3 shows an example of an early green that is granted after the bus arrives. The x-axis shows the position in the cycle which has a fixed length $C$ and begins with the green phase of length $G$.

The delay with no priority request (or when a request is denied) is as follows. In these expressions, $t_{a,rel}$ is specified relative to the beginning of the cycle. The conversion from absolute $t_a$ to relative $t_{a,rel}$ depends on the offset at the current signal, $f_i$: $t_{a,rel} = t_a - f_i - \lfloor \frac{t_a - f_i}{C} \rfloor C$.

$$D_n(t_{a,rel}) = \begin{cases} 0 & : t_{a,rel} \leq G \quad \text{arrives on green} \\ C - t_{a,rel} & : t_{a,rel} > G \quad \text{arrives on red} \end{cases}$$

The delay with a priority request is as follows:

$$D_p(t_{a,rel}) = \begin{cases} 0 & : t_{a,rel} \leq G & \text{arrives on green} \\ 0 & : t_{a,rel} \leq G + \frac{C-G}{2}, t_{a,rel} - r \leq G & \text{successful green extension} \\ C - t_{a,rel} & : t_{a,rel} \leq G + \frac{C-G}{2}, t_{a,rel} - r > G & \text{failed green extension} \\ 0 & : t_{a,rel} > G + \frac{C-G}{2}, r \geq Y_0 & \text{successful early green} \\ Y_0 - r & : t_{a,rel} > G + \frac{C-G}{2}, r < Y_0 & \text{late early green} \end{cases}$$

A graphical representation of the above expressions is given in Figure 3.4, with $r = 10$s and $Y_0 = 25$s. The 'early green constrained' line reflects the stated assumption that an early

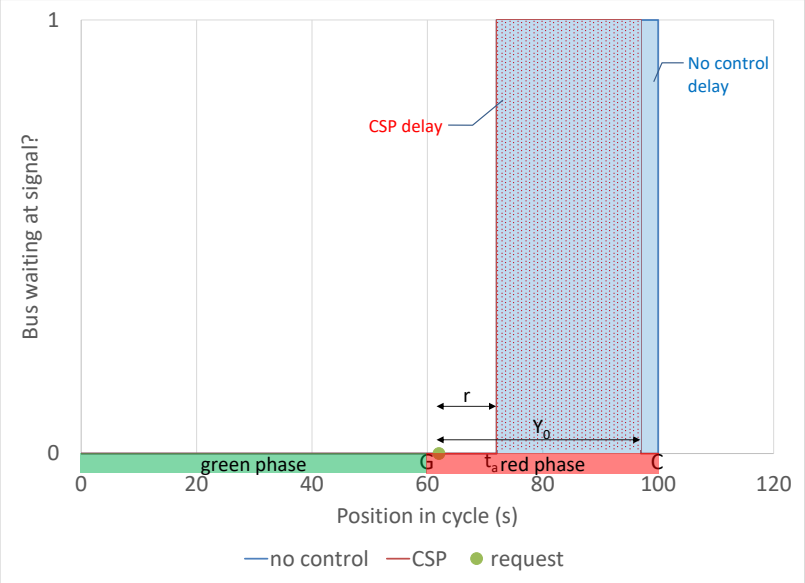



Figure 3.3: Depiction of advance notice and clear lag

green will not be granted before half of the red phase has elapsed. The 'early green' line does not contain this assumption. With the parameter values in this example, the half phase constraint is inactive but it might be relevant with a longer red phase. It is apparent that when the advance notice is short, there is only an advantage to sending a priority request in the first part of the red phase; in the latter part of the red phase, the pedestrian countdown is already underway and the request has no effect.

Because it was assumed that buses arrive randomly within the cycle, Figure 3.4 can be used to calculate the expected signal delay, and by extension the expected time savings when priority is requested. The expected signal delay with any control action is the area between that curve and the line bus departure=bus arrival divided by the cycle length. For no control, this expected delay for these signal timing parameters would be:

$$\begin{aligned} E[D_n] &= \frac{1}{C}\left[(\text{area under no control}) - (\text{area under dep.=arr.})\right] \\ &= \frac{1}{C}\left[\left(\frac{C^2}{2} + \frac{(C-G)^2}{2}\right) - \frac{C^2}{2}\right] \\ &= \frac{(C-G)^2}{2C} \\ &= \frac{(100-40)^2}{2\cdot 100} = 8 \text{ s} \end{aligned} \tag{3.2}$$

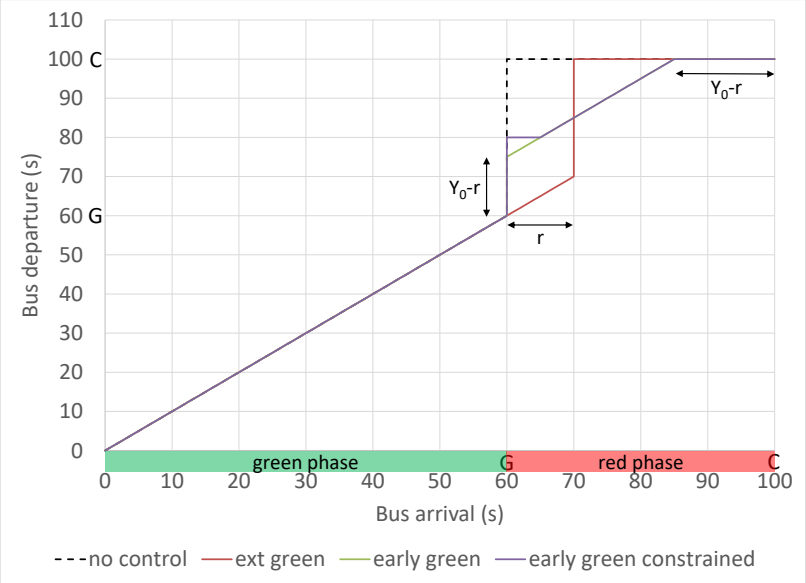



Figure 3.4: Graphical representation of control actions & time saved

The departure time with a priority request is the lower envelope of the *ext green* and *early green constrained* curves in Figure 3.4. The expected delay with a priority request for these signal timing parameters is:

$$\begin{aligned} E[D_p] &= \frac{1}{C}\left[(\text{area under lower env.}) - (\text{area under dep.=arr.})\right] \\ &= \frac{1}{C}\left[\left(\frac{C^2}{2} + \frac{(C-r-G)^2}{2} - \frac{(C-(Y_0-r)-G)^2}{2}\right) - \frac{C^2}{2}\right] \\ &= \frac{(C-r-G)^2}{2C} - \frac{(C-(Y_0-r)-G)^2}{2C} \\ &= \frac{(30)^2}{2\cdot 100} - \frac{(15)^2}{2\cdot 100} = 3.375 \text{ s} \end{aligned} \tag{3.3}$$

## 3.4 Variants

The previous section describes the basic form of conditional signal priority. Three variations are introduced here to provide a fuller comparison to the literature and to show that the rules and parameters of the previous section are the best available.

## Delay requests

As mentioned in the previous chapter, both Ma et al. (2010) and Chow and Li (2017) design CSP implementations for a particular corridor. The methods of both papers allow signal timing to be adjusted to delay, as well as speed up, buses. To provide a comparison to these works, we introduce a variant of CSP where buses can choose between sending a priority request, a delay request, or no request. Signals respond to priority requests with a green extension or early green, if the bus will arrive during the red phase, as described in the previous section. Delay requests work the same way, but in reverse. If the bus will arrive during the green phase, a delay request triggers a red extension or early red. To avoid excessive disruption to signal phases, red extensions are possible only in the first half of the green phase, and early reds only in the second half. When an early red phase is called, the red phase is of normal length. However, when $r < Y_0$, as is the case throughout this thesis, early red phases are not possible because the signal cannot be changed before the bus arrives. Unlike an early green, there is no point in changing the signal a few seconds after the bus arrives because it has already cleared the intersection.

## Threshold

As mentioned in the previous chapter, one property of the CSP strategy developed by Janos and Furth (2002) was a tolerance from schedule of 1 minute. Let $\delta$ represent the tolerance from schedule in seconds. Then CSP with a tolerance can be described by the following rule:

$$\begin{aligned} &\text{if } t_a - t_i > \delta: \\ &\qquad \text{send priority request} \\ &\text{else:} \\ &\qquad \text{no action} \end{aligned} \tag{3.4}$$

This control rule is the same as the one described in the previous section. In that case, $\delta = 0$, but other values can be tested in order to optimize the $\delta$ parameter.

## Predictive control

Let $t_d$ represent the bus' expected arrival time at the next downstream signal. This value is influenced by the control action taken at the present signal, which must belong to the set {priority request, no request, delay request}. A predictive control approach would replace the rules-based decision model, as represented by (3.4), with the following objective:

$$\min |t_d - t_{i+1}| \tag{3.5}$$

# Chapter 4

# Analysis, optimization, & evaluation

This chapter is organized as follows. Section 4.1 describes the scenarios of bus service for which the analysis, optimization, and evaluation is done. Section 4.2 discusses performance measurement and the tradeoffs involved in selecting control parameters for CSP, then introduces Brownian motion as an analytical model of bus motion. Brownian motion is used to develop formulas for the performance of CSP in steady state, and these formulas are verified in simulation. Section 4.3 uses the same simulation for comparative analysis, first looking at CSP, TSP, and no control and then to evaluate the variants proposed in Chapter 4. Section 4.4 is a discussion for the chapter, explaining what these results mean for bus operations.

## 4.1 Scenarios studied

The method of conditional signal priority has been described in the preceding sections in a general way, which can be applied to any bus route on any network. However, the choice of control parameters and the expected performance are likely to be different for different types of bus service. To derive some insights on the performance of conditional signal priority, we focuses on a subset of all bus routes: those with schedule-based service.

Schedule-based service is used here to refer to routes where adherence to schedule is the main operational priority. This is usually the case for infrequent bus routes, where passengers tend to consult the bus schedule and arrive at the station in time for a specific trip. Schedule adherence is important for these routes because an early departure could leave many passengers behind and a late departure translates to an equivalent amount of out-of-vehicle delay for schedule-conscious passengers.

Although the results in this thesis focus on schedule-based service, the method can be applied to all types of bus routes, including those without schedules. With any bus service, it is possible to make a distinction between a printed schedule, which is communicated to passengers, and a virtual schedule, which is only used for bus operations and is not shared with the public. In schedule-based service, it might be desirable to have a virtual schedule that is slightly later than the printed schedule by a fixed amount. This practice would ensure

that passengers are rarely left behind by an early bus, since buses could oscillate around the virtual schedule while remaining on-time or slightly late with respect to the printed schedule. Using this terminology of printed and virtual schedules, a system without a schedule is one that lacks a printed schedule. Instead, passengers may be informed of the typical headway, or be shown real-time information about upcoming arrivals. Such systems can still have a virtual schedule, and there is much greater flexibility to adapt the virtual schedule when it is not connected to a printed schedule. Each bus could be given a virtual schedule when it is dispatched on a task, and the downstream portions of this schedule could be adjusted later on based on the real-time position of other buses.

Because the focus is on schedule-based service, the pairing effect will be ignored. The pairing effect refers to the phenomenon where the schedule deviation of the bus in front affects the number of passengers encountered by the selected bus, eventually leading to bus bunching. As mentioned in the previous chapter, this effect is less important for schedule-based services because most passengers time their arrival to the schedule. In any system, there may be some passengers who arrive randomly without consulting the schedule. As long as these people are relatively small in number, the pairing effect will remain negligible.

In this thesis, it is assumed that buses are not running any form of schedule control except for the strategies of transit signal priority and conditional signal priority as specified. In schedule-based service, it is a standard practice to run a form of schedule control, where drivers compare their position to the schedule at certain time points and hold if they are early. Since US transit systems have a large number of optional stations, the time points are typically placed at major intersections and transfer points that are unlikely to be skipped. These time points are also where slack time can be inserted into the schedule. Schedule control and slack time are useful for returning early buses to schedule and allowing late buses to make up a small amount of delay. The decision not to use them in the model means that we are focusing on the worst case. This assumption is made in order to isolate the effects of transit signal priority and conditional signal priority.

## 4.2 Analysis and verification

Root Mean Square deviation (RMS) is used as the main performance measure. The reason why is related to the distinction between a virtual and printed schedule. Consider, for example, one control strategy which stabilizes buses at a mean of 30 seconds behind the virtual schedule with large variance and a second control strategy which stabilizes buses at a mean of 1 minute behind the virtual schedule with small variance. Which strategy is better? The first minimizes schedule deviation, but in practice the position of the bus with respect to the virtual schedule is unimportant. The printed schedule, if there is one, can be set based on where the bus stabilizes. By doing so, we assume that the expected schedule deviation can be set to 0, $E(\varepsilon) = 0$, in which case the root mean square deviation is equivalent to the standard deviation of travel time, $RMS \equiv \sigma$. If there is not a printed schedule, then stable headways are all that matters. In both cases, the smaller RMS is better for passengers.

Generally speaking, the goal of conditional signal priority is to stabilize bus trajectories. In quantitative terms, this is done by minimizing RMS deviation.

Why focus on RMS deviation? There are several reasons, first that it is directly observable by the bus operator. Good performance metrics are easily observable and either represent, or are correlated with, important outcomes. In schedule-based service, RMS deviation correlates with out-of-vehicle delay at origin and transfer stations but is easier to observe. Out-of-vehicle delay depends on the number of passengers, which station they board at, and when they arrived at the station. RMS deviation, on the other hand, is an operator-side metric that can be calculated from vehicle position data that the operator likely already possesses.

## Tradeoffs

The selection of control parameters, which for conditional signal priority means the design of the virtual schedule, involves two main tradeoffs.

The first tradeoff is between stability and speed. It is expected that the fastest possible schedule is also the least stable. Any delay, once incurred, is unrecoverable with signal priority because the control cannot sustain a faster average speed. This property is an important shortcoming of TSP as pointed out in Chapter 1. In current practice, transit agencies use informal measures like boarding limits, stop skipping, and short turns to return to schedule in these situations. However, these measures involve another tradeoff: delaying some passengers by a headway in order to make up time. Returning to the issue of stability vs. speed, one expects that many schedules are stable, but that some may take longer to reach a steady state. Properties at steady state are important for the theory and will be examined later in this section. In practice, all bus routes are finite so the time to stabilize and the pre-steady state conditions are also important. Finally, there are externalities related to the speed of the schedule. A slower speed implies fewer priority requests which means a smaller impact on other traffic and pedestrians. These externalities may politicize the decision and lead to a different virtual schedule than the one that would be preferred for stability alone.

The second tradeoff is between minimizing variance and delay recovery. It will be shown in Chapter 5 that the rate of delay recovery is determined by the difference in average speed between TSP and the virtual schedule. The tradeoff, therefore, is between stabilizing a bus route and providing robustness to large delays. Different objectives may be appropriate for different bus routes.

## Brownian motion background

One tool which will be used in the analysis of conditional signal priority is Brownian motion. Brownian motion (also known as the Wiener process) is a continuous-time stochastic process which is widely used in applied mathematics. It is used in this thesis both to categorize processes and to support claims about their properties.

The discovery of the physical phenomenon is usually credited to Brown (1828). The mathematical description of the process originates with Thiele (1880) and was further developed by the work of Bachelier (1900) in finance and Einstein (1905) and Von Smoluchowski (1906) in physics. Wiener (1923) rigorously described the existence and construction of Brownian motion, hence the process is sometimes referred to using his name.

Brownian motion, $B$, is a stochastic process $\{B(x) : 0 \leq x < \infty\}$ with the following properties:

1. B(0)=0.
2. Independent increments: future increments $B(x+z) - B(x), z \geq 0$, are independent of past values $B(y), y \leq x$.
3. Gaussian increments: $B(x+z) - B(x) \sim N(mz, \sigma z)$.
4. Continuous paths: $B(x)$ is continuous in $x$.

In item 3, $m$ is the drift and $\sigma$ the variance rate. The process is *state dependent* if $m$ and/or $\sigma$ depend on the state $B(x)$. A special case is the unitary Brownian process, where $m = 0$ and $\sigma = 1$.

If $m$ is constant and different from 0, the process $B$ will go to $\infty$ or $-\infty$ (depending on the sign of $m$) as $x \to \infty$. Let $m > 0$, then it is inevitable that the process $B$ will pass through a state $\varepsilon$, $0 < \varepsilon < \infty$, at least once. Let $x_0$ be the value of $x$ at first passage. This $x_0$ is a random variable which follows an inverse Gaussian distribution:

$$x_0 = \inf\{x > 0 | B(x) = \varepsilon\} \sim IG\left(\frac{\varepsilon}{m}, \frac{\varepsilon^2}{\sigma^2}\right). \tag{4.1}$$

The inverse gaussian distribution is defined by the parameters $\mu = \frac{\varepsilon}{m}$ and $\lambda = \frac{\varepsilon^2}{\sigma^2}$. The expected value of an inverse gaussian and therefore the first passage $x_0$ is:

$$E(x_0) = \mu = \frac{\varepsilon}{m}, \tag{4.2}$$

and its variance is:

$$Var(x_0) = \frac{\mu^3}{\lambda} = \frac{\varepsilon\sigma^2}{m^3}. \tag{4.3}$$

## Brownian motion analysis

We propose that the schedule deviation of a bus controlled by conditional signal priority can be represented by a state-dependent Brownian motion with drift, as described in the previous section. Schedule deviation starts at 0 at distance $x = 0$, provided the bus is initially dispatched on schedule. Schedule-based bus service has independent increments since the pairing effect can be neglected. Increments represent the change in schedule deviation due

to variations in signal delay, traffic effects, and dwell time. It is reasonable to approximate these effects with a normal distribution. Finally, distance, bus motion, and the schedule function defined in Section 3.2 are all continuous, so schedule deviation is also continuous.

Recall that $T_u$, $T_s$, and $T_c$ are the average paces of TSP, CSP, and no control in seconds. Let $\varepsilon$ represent schedule deviation, $m = T_c - T_u$, $m_+ = T_c - T_s$, $m_- = T_s - T_u$, $\gamma = m_+/m_-$, and $\sigma^2$ represent variance of time per stop.

Define schedule deviation as a stochastic process $\varepsilon(x)$. This process is a Brownian motion with state-dependent drift, $m$, and variance rate, $\sigma$. Drift takes on two values, $m_+ > 0$ when $\varepsilon(x) \leq 0$ (i.e. the bus is early or on time), and $m_- > 0$ when $\varepsilon(x) > 0$ (i.e. the bus is late):

$$d\varepsilon(x) = \left\{ \begin{array}{ll} m_+ & \varepsilon(x) \leq 0 \\ -m_- & \varepsilon(x) > 0 \end{array} \right\} dx + \sigma dx, \quad \varepsilon(0) = 0 \tag{4.4}$$

Constraining $m_+$ and $m_-$ to be positive ensures that $\varepsilon(x)$ will not diverge to $-\infty$ or $+\infty$, respectively. The expectation and variance of schedule deviation were developed in Daganzo (2017). First, it is assumed that positive and negative schedule deviations are equally likely when $\varepsilon \approx 0$. Then, positive and negative occupation times depend on $\gamma = m_+/m_-$, the ratio of the drift values. The fraction of time that $\varepsilon(x)$ is positive is $\frac{\frac{1}{m_+}}{\frac{1}{m_+}+\frac{1}{m_-}} = \frac{1}{\gamma}$ and the fraction of time that $\varepsilon(x)$ is negative is $\frac{\frac{1}{m_-}}{\frac{1}{m_+}+\frac{1}{m_-}} = \gamma$. Finally, because the drift on each side of 0 is constant, the mean positive and mean negative deviations are half the initial random deviation multiplied by the respective occupation time. The formulas are:

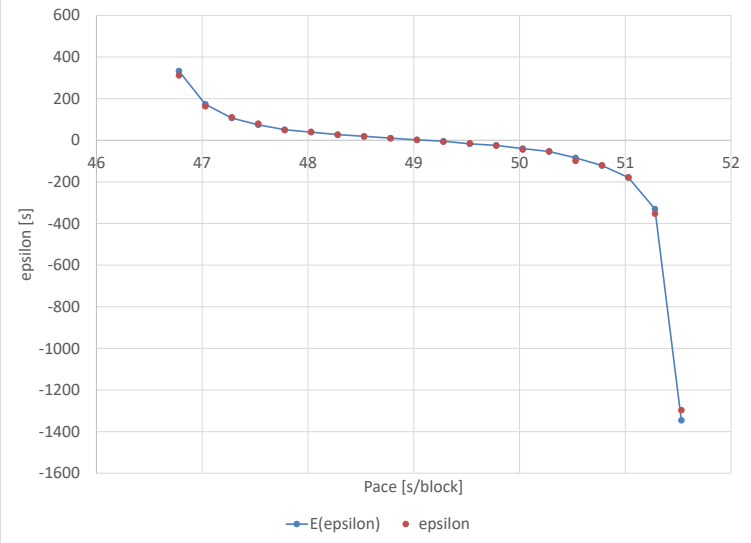
$$E(\varepsilon) = \frac{\sigma^2}{m}\left(\frac{\gamma^2 - 1}{2\gamma}\right) \tag{4.5}$$

and

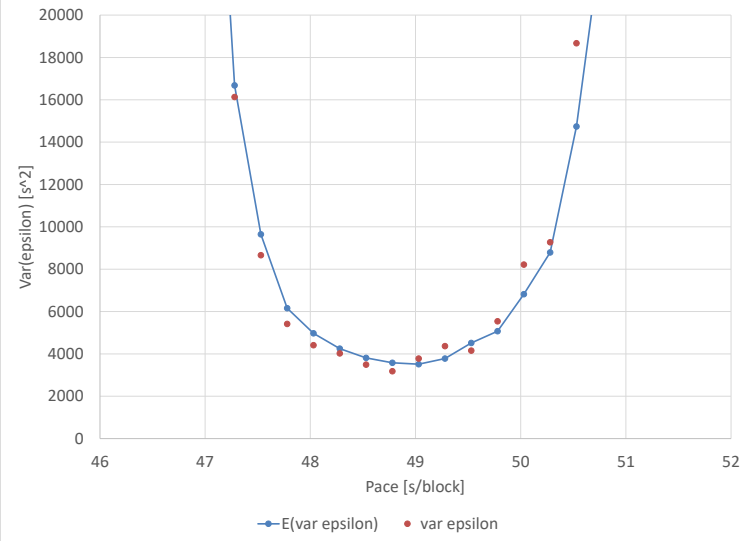
$$Var(\varepsilon) = \frac{\sigma^4}{m^2}\left(\frac{(1+\gamma)^2(1+\gamma^2)}{4\gamma^2}\right). \tag{4.6}$$

These equations can be tested by simulating bus motion for different values of $T_s$. The simulation represents the network model described in the previous chapter. There are 1,000 draws and each draw represents the passage of one bus on a route over a given number of signals (i.e. blocks). The parameters in the simulation are: signal spacing $s = 0.25$mi, bus route spacing $b = 0.25$mi, cycle length $C = 100$s, green phase $G = 60$s, clear lag $Y_0 = 20$s, bus cruising speed $v = 30$mph, station spacing $x = 0.25$mi, dwell time $[N(12.3, 13.5)]^+$s, conflicting bus route headway $h_i = 600$s $\forall i$, advance notice $r = 10$s. Because signal delay contributes to travel time variance, $\sigma$ varies with $T_s$ so it is estimated in the simulation. For each value of $T_s$, there is a warmup period that lasts until schedule deviation stabilizes. The warmup period is 9,000 signals for $46.78 \leq T_s \leq 51.28$s/block and 90,000 signals for $T_s = 51.53$s/block. Schedule deviation and variance are then computed from the next 1,000 signals (10,000 for $T_s = 51.53$s/block). The longer warmup time for $T_s = 51.53$s/block was necessary because $\gamma \to 0$ as $T_s \to T_c$, meaning that $E(\varepsilon) \to \infty$.

The simulation results are shown in Figure 4.1. The experimental schedule deviation and variance (in red) are a very close match to the predictions from the above equations (in blue), confirming the validity of the Brownian motion approximation. As predicted by (4.5), the expected value of schedule deviation goes to 0 when the two drifts are balanced, which occurs at approximately $T_s = 49.03$s/block.




(a) Predicted and actual schedule deviation




(b) Predicted and actual variance of schedule deviation

Figure 4.1: Verification of Brownian motion formulas

## 4.3 Comparison

### CSP compared to TSP and no control

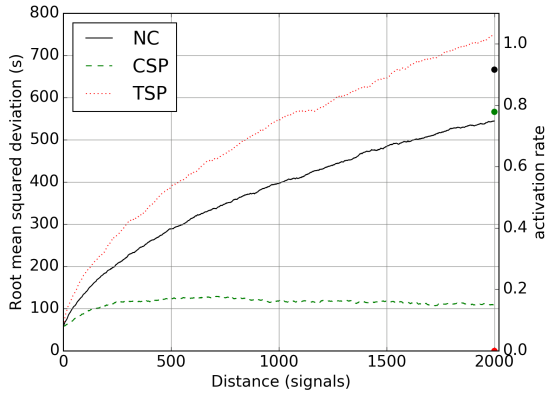
The previous section shows that conditional signal priority guides buses to a steady state where the expectation and variance of schedule deviation can be accurately predicted by Brownian motion. This result is important theoretically, but for practical considerations it is important to know how long it takes to reach steady state, and how the performance of CSP compares to TSP and no control. The simulations in this section use the same parameters listed in the previous section unless stated otherwise.

The first question is how long it takes CSP to reach a steady state. It is expected that the answer will depend on the pace (which affects the symmetry of drift about the schedule) and on the initial conditions. Regarding initial conditions, transit agencies strive to dispatch buses on schedule but the precision may only be to the nearest minute. Some agencies, e.g. Swiss Federal Railways and Japan Railways, do achieve dispatch precision on the order of seconds, but this is rare. To reflect the real-world situation, dispatches are randomized with a distribution of $U(-C/2, C/2)$. Buses are also impacted by the offsets between signals. To eliminate this confounding influence, a random set of offsets is drawn from $U(0, C)$ in each simulation draw.

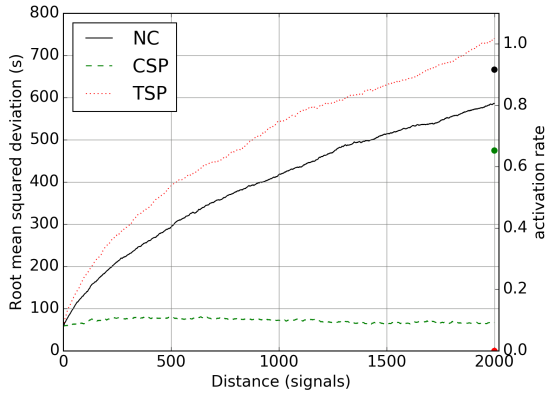
Figure 4.2 contains four plots, each showing CSP with a different pace compared to TSP and no control. The primary y-axis is the root mean square schedule deviation. The secondary y-axis shows the realized activation rate, and the realized commercial speed of CSP is shown above the plot.

Figure 4.2(a), reaches steady state after about 600 signals, Figure 4.2(b) after about 300 signals, Figure 4.2(c) almost immediately, and Figure 4.2(d) around 800 signals. With the exception of Figure 4.2(c), these distances to steady state (75 mi and up) are significantly longer than the length of most bus routes. Even so, the RMS deviation with TSP and no control grows without bound. CSP has a lower RMS deviation than TSP or no control both before and after reaching steady state, and this is true for all four values of pace.

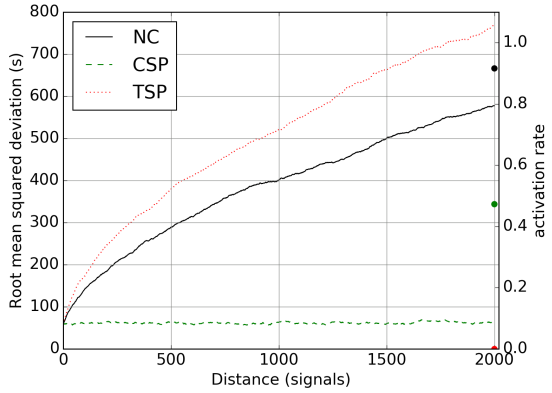
The activation rate, shown on the secondary y-axis, represents the disruption that signal priority inflicts on other vehicles. Note that the activation rate of TSP is only 0.92. It is less than 1.0 because priority requests are denied when the signal is already serving a conflicting request. The method of conflict resolution is described in Section 3.1. The activation rate of CSP is inversely related to the pace that is used. In Figure 4.2(a) it is 0.78, in Figure 4.2(b) 0.65, in Figure 4.2(c) 0.47, and in Figure 4.2(d) 0.13. The sharp drop in activations as $T_s$ increases suggests that a slower pace could be used if the disruption to other traffic is a significant concern. It often is a concern, even with the longer headways of schedule-based service, there could be a request every other cycle at an intersection with two bus routes using TSP.



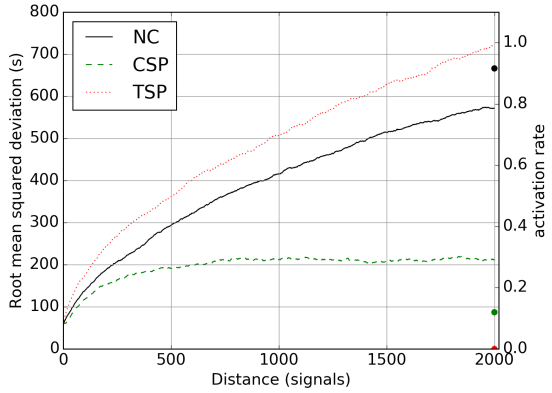

(a) Root mean squared schedule deviation, $T_s = 47.28$s/block




(b) Root mean squared schedule deviation, $T_s = 48.03$s/block




(c) Root mean squared schedule deviation, $T_s = 49.03$s/block




(d) Root mean squared schedule deviation, $T_s = 51.03$s/block

Figure 4.2: Root mean squared schedule deviation for different CSP paces

## Variants: Delay Requests and Predictive Control

The two previous sections covered the results for the basic form of conditional signal priority. This section and the next evaluate the variants described in Section 3.3. Two variants are introduced in this section: CSP with delay requests as described in Section 3.3, and CSP with delay requests and predictive control. Predictive control changes the way that requests are chosen. The other forms of CSP have a decision rule that is based on the current state of the bus, $\varepsilon(x)$. Predictive control projects the value of $\varepsilon(x + s)$ for each possible control action and selects the one closest to 0. It could also be applied to the basic form of CSP, but for the sake of comparison its impact is easiest to see when it can choose between a priority request, delay request, or no request.

To see the impact of these variants, we can plot RMS deviation for the same range of

$T_s$ values shown in Figure 4.1. This plot, Figure 4.3, shows that the two variants produce the same results as CSP for $T_s < 48$ s/block. This result makes sense, at faster paces drift is asymmetric and buses are much more likely to be late than early. The two variants have an additional control action, the delay request, but few opportunities to use it. From $48 \leq T_s \leq 49$ s/block, buses still run late but the occasional delay request is effective at lowering the RMS deviation. For $T_s > 49$ s/block, drift is asymmetric in the other direction and buses run early. In this case, the two variants send more delay requests than priority requests. For all $T_s$ values, predictive control reduces RMS deviation by suppressing some requests that would overcorrect for small schedule deviations. Overcorrection is more of a problem for delay requests than priority requests, as shown by the widening gap when $T_s > 49$ s/block.

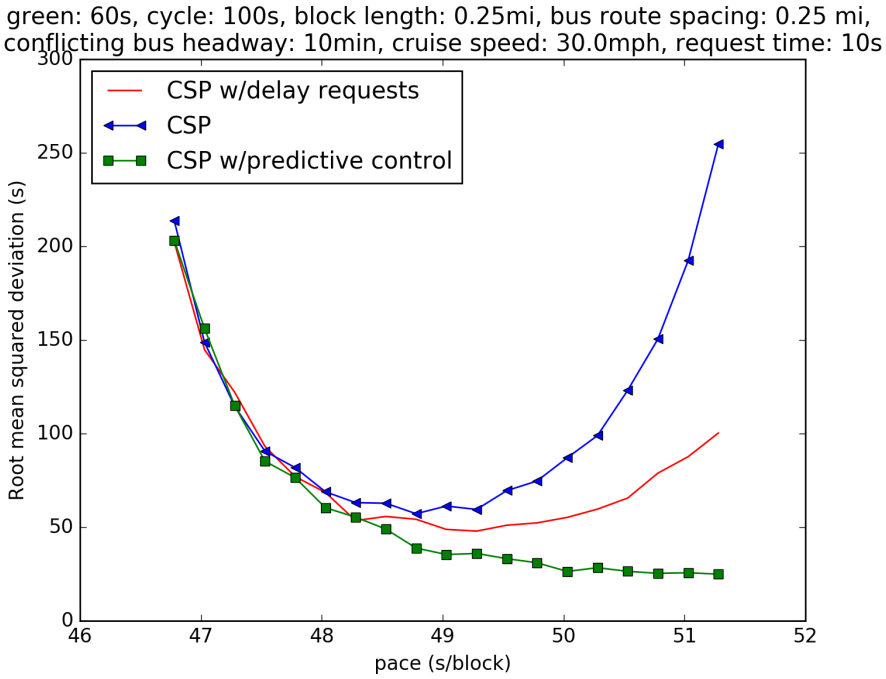
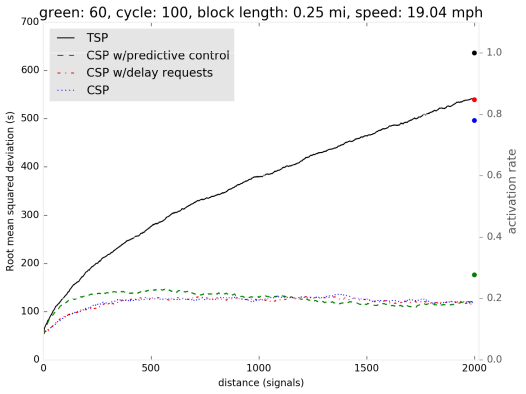


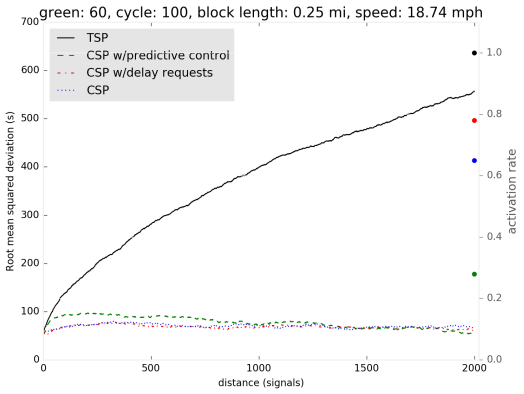

Figure 4.3: Root mean squared schedule deviation for different CSP paces

As shown in Figure 4.3, delay requests are effective at reducing RMS deviation in certain cases. Figure 4.4 shows the variants on the same plots as Figure 4.2. A clear advantage of CSP with predictive control in Figure 4.4(d) is not only its low RMS deviation but its fast stabilization time.

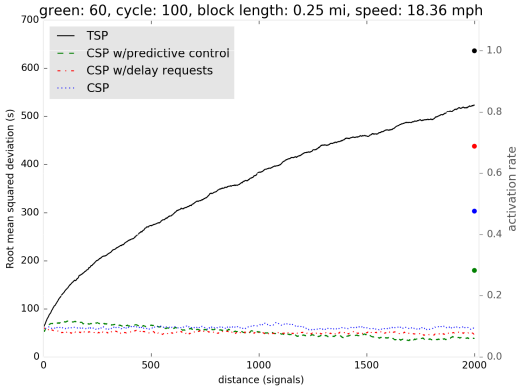
However, the problem with delay requests is not bus performance but their impact on signal timing and other vehicles. Figure 4.4 shows activation rates on the secondary y-axis. By suppressing some unnecessary requests, CSP with predictive control has a consistent activation rate of 28%. CSP and CSP with delay requests have activation rates that vary with $T_s$, but CSP is always lower. For $T_s = 51.03$ s/block, CSP is the lowest overall.



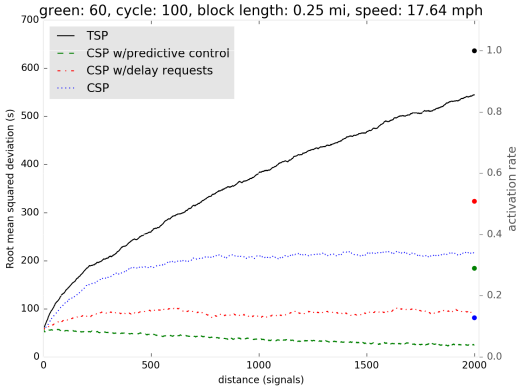

(a) Root mean squared schedule deviation, $T_s = 47.28$s/block




(b) Root mean squared schedule deviation, $T_s = 48.03$s/block



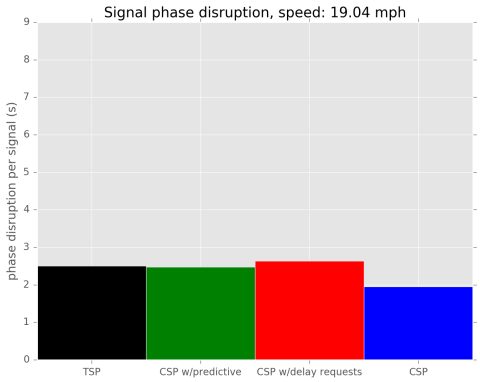

(c) Root mean squared schedule deviation, $T_s = 49.03$s/block



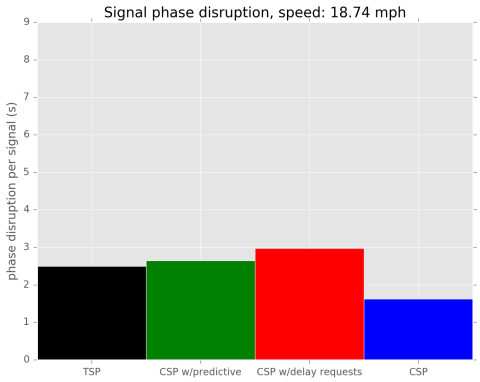

(d) Root mean squared schedule deviation, $T_s = 51.03$s/block

Figure 4.4: Root mean squared schedule deviation for different CSP paces

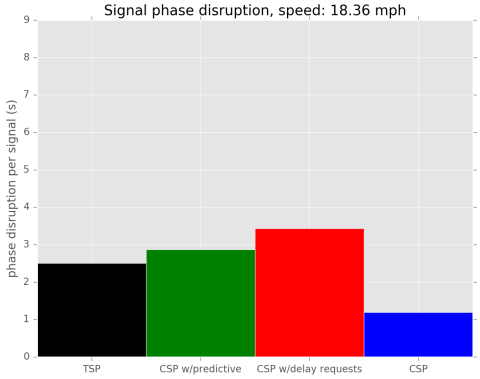
Figure 4.5 shows the disruption that the four control strategies cause to normal signal timing. Phase disruption is defined as the maximum delay to a vehicle that would normally have gone through the intersection on green. Using this definition, a 10s green extension counts as 10s, while an early red phase counts as $C - G$. As Figure 4.5 shows, delay requests are hugely disruptive. CSP has fewer activations and smaller phase disruption than TSP in all cases, while both variants that include delay requests cause at least as much phase disruption as TSP, even when they have many fewer activations. Phase disruption increases the average delay at the intersection, even if lost green time is compensated in the next cycle. If vehicles arrive randomly, the argument is similar to headway-based bus service, where regular headways minimize waiting time. If signals have been timed to provide some progression, phase disruption may result in an entire platoon arriving during red instead of green.



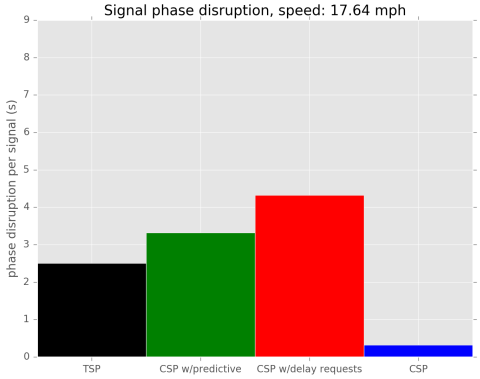

(a) Phase disruption, $T_s = 47.28$s/block




(b) Phase disruption, $T_s = 48.03$s/block




(c) Phase disruption, $T_s = 49.03$s/block




(d) Phase disruption, $T_s = 51.03$s/block

Figure 4.5: Phase disruption of different control variants

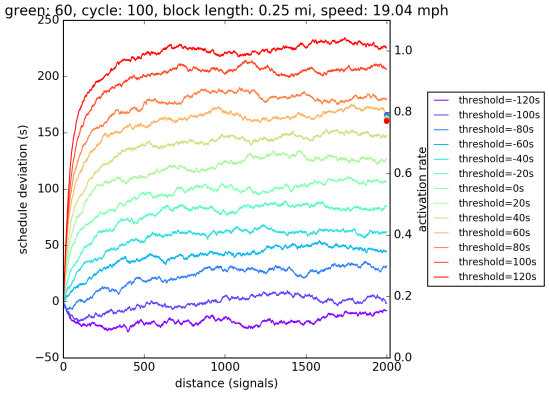
Another way to interpret the results is to say that Figure 4.3 shows the benefits of applying some form of schedule control that holds early buses until their scheduled departure time. Traditionally, this has been the responsibility of bus drivers, and the holding time is applied at certain stations known as time points. Viewed in this context, delay requests are an attempt to automate a responsibility of bus drivers by giving it to traffic signals instead. This effort does not seem to be worthwhile.

## Variants: Threshold

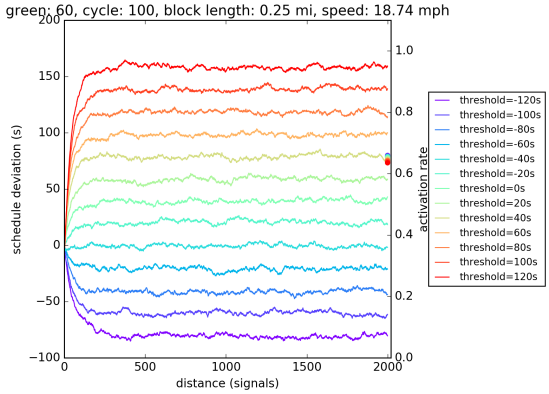
This section explores the use of thresholds, as described in Section 3.3. Figure 4.6 and Figure 4.7 plot schedule deviation and standard deviation of schedule deviation, respectively, for the same selection of four $T_s$ values used in the previous two sections. These plots include thresholds in the range [-120,120] seconds at multiples of 20 seconds.

Previously, $T_s = 49.03$ s/block was shown to have an expected schedule deviation of

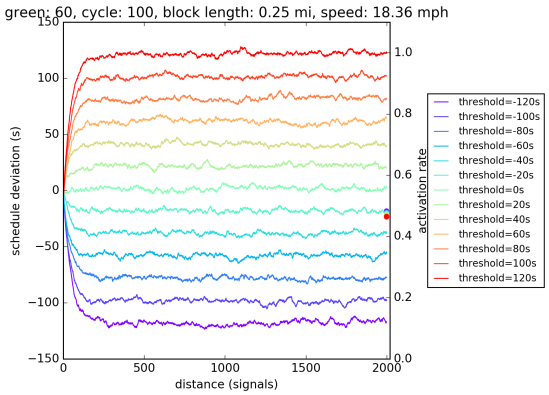
0. Figure 4.6(c) confirms that the average schedule deviation is 0 in this case, and shows that any nonzero value of threshold translates directly to a schedule deviation of the same magnitude. The other subplots of Figure 4.6 show that for any other value of $T_s$, a nonzero threshold value can correct for the schedule deviation that results from asymmetric drift. This result is useful for situations where a schedule already exists and the purpose is to construct a control strategy that maintains it.



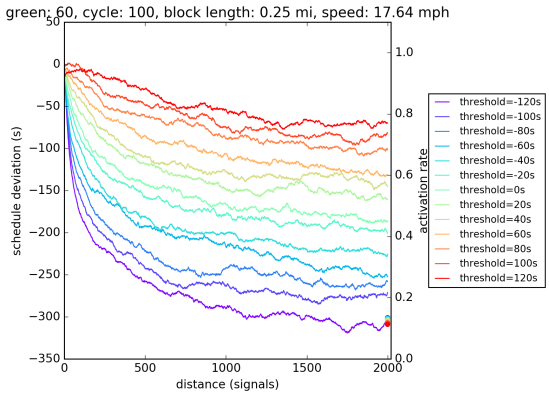

(a) Schedule deviation with thresholds, $T_s =$ 47.28s/block




(b) Schedule deviation with thresholds, $T_s =$ 48.03s/block




(c) Schedule deviation with thresholds, $T_s =$ 49.03s/block

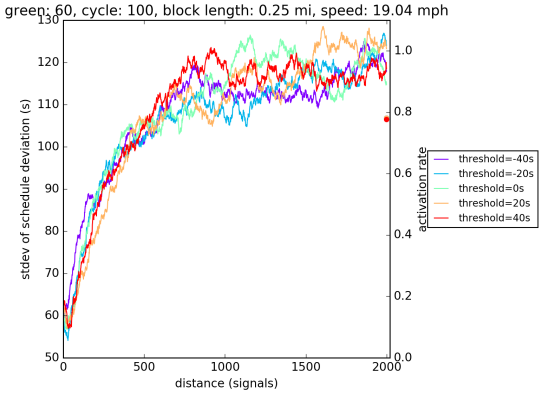



(d) Schedule deviation with thresholds, $T_s =$ 51.03s/block

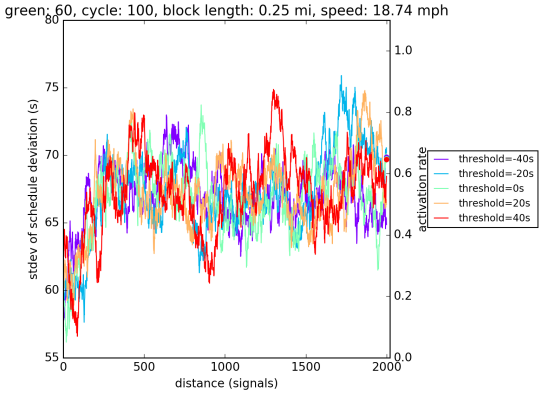
Figure 4.6: Schedule deviation with thresholds for different CSP paces

Figure 4.7 shows that the standard deviation of schedule deviation is similar for all threshold values. The implication is that thresholds are unimportant in steady state. It is merely a question of whether the offset between virtual and printed schedules is endogenous or exogenous to the CSP strategy. However, the stabilization period observed in Figure 4.6 is relevant for real-world operations. In these simulations, buses were dispatched based on the virtual schedule, with some scatter as mentioned earlier. If a nonzero threshold is used,

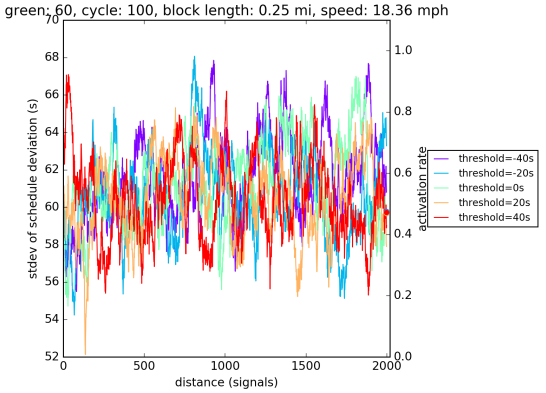
then the stabilization period could be significantly shortened by dispatching buses based on the virtual schedule plus the threshold.



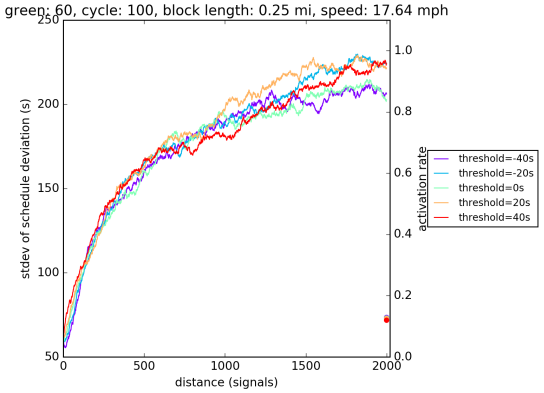

(a) Standard deviation of schedule deviation with thresholds, $T_s = 47.28$s/block




(b) Standard deviation of schedule deviation with thresholds, $T_s = 48.03$s/block




(c) Standard deviation of schedule deviation with thresholds, $T_s = 49.03$s/block




(d) Standard deviation of schedule deviation with thresholds, $T_s = 51.03$s/block

Figure 4.7: Standard deviation of schedule deviation with thresholds for different CSP paces

## 4.4 Macroscopic view and benefits

The preceding results have shown that the proposed method of conditional signal priority can stabilize buses, guiding them to a steady state where bus trajectory is a Brownian motion. In this steady state, the expected schedule deviation and its variance are given by Brownian motion formulas. This means that root mean squared schedule deviation is bounded with CSP, while it grows without bound with TSP or no control.

What do these theoretical findings mean for implementation and benefits of this strategy in practice? First, CSP is widely applicable because it has few technological requirements. Since buses make their own priority decisions and the vehicle location/communications equipment is also kept on the bus, the only requirement for fixed infrastructure is a signal capable of receiving requests. The simulations in this thesis assumed fixed-time signals, which represent 65% of the traffic signals in San Francisco as of July 1, 2017. This percentage is probably typical of large, established cities. There is a trend of cities replacing older, fixed-time signals with actuated signals. This effort may benefit transit signal priority by eliminating one complaint: that it disrupts signal progression for cars.

One of the basic properties of CSP is that it does not rely on receiving priority at every signal. This property has implications for applicability as well as integration with other control strategies.

The minimum requirement for implementing CSP is that the bus route has at least one signal where priority can be requested. It is likely that most bus routes can meet this requirement. If only a few signals can receive priority requests, the strategy will be quite constrained but it could still improve schedule adherence relative to no control.

The previous statement remains true if the limiting factor is not technological capability but flexibility in signal timing. Bus routes tend to be located in urban areas where there may be other control strategies in use such as perimeter control or backward offsets. A single priority activation is not critical to CSP, so it may be best to deny a priority request that would undermine another control strategy.

The next chapter will address delay recovery with CSP. Recovery works the same way whether the delay results from traffic interactions, dwell time variability, or if it accumulates due to priority being unavailable on a segment of the route.

The benefits of CSP can be estimated using three different metrics. One is expected travel time, and on this measure CSP is faster than no control but slower than TSP. A second is out-of-vehicle delay. In schedule-based service, this is the schedule deviation when the bus is late, or an entire headway if an early bus leaves passengers behind. Here CSP performs better than TSP or no control. Finally, the planning time (e.g. 95th percentile travel time) is important to people who are trying to arrive at their destination on time. This correlates with RMS deviation and CSP again performs better than the others.

# Chapter 5

# Conditional Signal Priority to improve transfer times

## 5.1 How CSP enables transfer coordination

This chapter focuses on how conditional signal priority (CSP) can be used to improve transfer times on a network of two or more intersecting bus routes. It will be shown in the next section that CSP-controlled buses have the ability to recover from delays of any size. In the previous chapter, Brownian motion was used to develop formulas for the expectation and variance of schedule deviation in steady state. In this chapter, the same mathematical model will be used to develop formulas for the expectation and variance of distance to recovery after a delay.

The ability to recover delays in a predictable way enables a strategy of transfer coordination, which will also be introduced in this chapter. Transfer coordination has the potential to shorten transit trips and to make travel times by transit more reliable. As an added benefit, the time saved is out-of-vehicle waiting time which passengers dislike more than any other component of their trip. However, transfer coordination is only worth doing if the delay from holding for additional passengers can be made up. Otherwise the downstream effects are too large and the hold cannot be justified. Previous papers on transfer coordination have not found a satisfactory way to make up the holding delay without harming other passengers. Conditional signal priority may be the answer to that problem.

CSP also helps transfer coordination by keeping buses closer to the schedule. Any static transfer coordination that is designed into the bus schedule will work better when schedule deviations are small. Adaptive transfer coordination also works best when bus arrivals are already close together. Combined, CSP and transfer coordination can improve transit travel time by reducing late buses and missed transfers.

The metric that will be used in transfer coordination is passenger time. This objective is complementary to the CSP goal of reducing root mean squared schedule deviation because RMS deviation is correlated with passenger delay. CSP uses an operator-side objective to

avoid having to count passengers. In transfer coordination, passenger time can be used directly because passenger counts are necessary to implement the strategy. For transferring and downstream passengers, the time in question is spent waiting at the station. For passengers already on the bus, the delay is measured when they get off the bus. This assumption is reasonable because if holding delay can be recovered quickly, passengers may not notice or care. They may not notice because large bus stations, where many transfers happen, already have longer dwell times than an isolated station served by a single route. Large stations tend to have more passenger movements (boardings and alightings) and are likely to be control points if schedule control is used. They may not care because in-vehicle time is the least onerous part of a transit trip. More importantly, transportation is (usually) not a goal in itself. For some trips, passengers care about arriving on time, for others the goal is simply to get there quickly. Measuring delay on alighting is consistent with getting passengers to their destinations.

The strategy of transfer coordination presented here is an opportunistic strategy which considers all transfers as potential candidates for coordination. Opportunistic action is easier than planning in advance. For one, people may use the bus system in a way that is different from the original design. Many bus routes in US cities follow old streetcar routes, yet land use and travel patterns have changed immensely in the last 120 years. Opportunism can identify and benefit transfers that are not planned in the schedule. Additionally, an opportunistic strategy continues to work in disrupted conditions when planned transfers become meaningless. The main benefit of this type of strategy is that it is simple and can be applied everywhere. The control can run in the background at all transfer points. There is no cost (aside from the minimal computation cost) if it does not intervene, and action is taken only when worthwhile.

The rest of this chapter is organized as follows. Section 5.2 covers the theory of delay recovery with CSP, including the formulas obtained from Brownian motion and their verification in simulation. Section 5.3 introduces the strategy of transfer coordination, first for a simple case of one connecting bus with certain information, then the simple case with uncertain information, and finally a general case with multiple connecting buses. In Section 5.4, a case study is developed from observations taken at a transfer point in Oakland and simulation results are shown to offer insights about the expected performance in other situations. The chapter concludes with a discussion in Section 5.5 of the benefits that can be expected from a large-scale implementation of conditional signal priority with transfer coordination.

## 5.2 Delay Recovery

Let $\varepsilon_0$ be the initial lateness of a bus. It is irrelevant whether this lateness came from a point delay or was accumulated over time. Then let $x_0$ be the distance to full recovery, where full recovery is defined as the intersection of this bus' trajectory with the schedule. The expectation of distance to full recovery is given by the transient of Brownian motion.

Recall the definition of the stochastic process $\varepsilon(x)$ in Equation (4.4) as a Brownian motion with two-valued drift:

$$d\varepsilon(x) = \left\{ \begin{array}{ll} m_+ & \varepsilon(x) \leq 0 \\ -m_- & \varepsilon(x) > 0 \end{array} \right\} dx + \sigma dx, \quad \varepsilon(0) = 0.$$

Both values of drift are positive as defined and pull the process back to 0.

The literature definition of first passage was given in Section 4.2. These expressions describes the expectation and variance of first passage, $x_0$, for a Brownian motion $B$ with drift $m$ starting at $B(0) = 0$. The transient of interest for buses, representing the recovery of delay, is a translation and reflection of the literature definition. Suppose that a bus is behind schedule by some amount $\varepsilon_0$. Let us reset $x$ so that $\varepsilon(0) = \varepsilon_0$. Since $\varepsilon(0) > 0$, the drift is $-m_-$, at least until the process dips below 0. Since we are interested in the first passage through 0, $x_0$, the drift is always $-m_-$.

The first passage through 0 of a process starting at $\varepsilon(0) = \varepsilon_0$ with drift $-m_-$ is analogous to the literature definition of the first passage through $\varepsilon_0$ of a process starting at $\varepsilon(0) = 0$ with drift $m_-$. Both can be represented by an inverse Gaussian:

$$x_0 = \inf\{x > 0 | \varepsilon(x) = 0\} \sim IG\left(\frac{\varepsilon_0}{m_-}, \frac{\varepsilon_0^2}{\sigma^2}\right). \tag{5.1}$$

Therefore, the expected value of $x_0$ is:

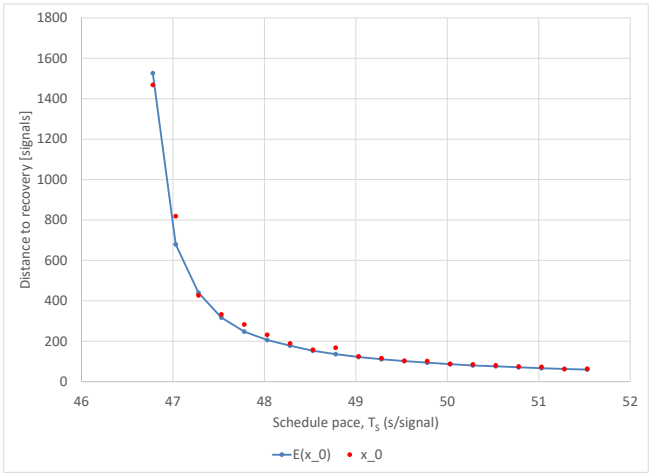
$$E(x_0) = \mu = \frac{\varepsilon_0}{m_-}, \tag{5.2}$$

and its variance is:

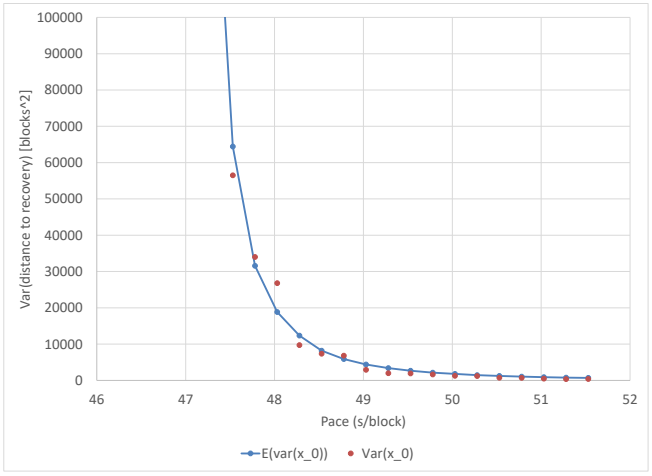
$$Var(x_0) = \frac{\mu^3}{\lambda} = \frac{\varepsilon_0 \sigma^2}{m_-^3}. \tag{5.3}$$

The results in Chapter 4 showed that conditional signal priority bounds deviations from schedule under normal operating conditions. This is already a useful result. However, buses sometimes experience major disturbances which impose a large point delay, and it is important to test the response of the control strategy to these events. The formulas above assume that the control will impose a drift of $-m_-$ regardless of the size of the delay. This property means that delays of any size can be recovered and that CSP has no escape problem.

As in Chapter 4, these formulas can be verified in simulation. Figure 5.1 shows the results of a simulation where a point delay of 300 seconds was imposed on all trips at the origin. After the starting delay, each bus proceeds along the route as before, with the same parameters and rules for requesting priority. The simulation consists of 100 draws, where each draw is the passage of 1 bus over a 100,000 signals. The long length of the simulation is necessary because some trajectories asymptotically approach the schedule and take a long distance to cross it. The distance to recovery is not calculated until the bus intersects the schedule. Any draw where the bus has not met this condition for recovery by the end of the simulation is thrown out.




(a) Predicted and actual distance to recovery for different schedule paces




(b) Predicted and actual variance of distance to recovery

Figure 5.1: Predicted and actual delay recovery

Figure 5.2 fixes the pace at 48.28 s/block and repeats the simulation for different values of $\varepsilon_0$, the starting delay. The length of each draw is reduced to 10,000 signals for this plot since only the mean is being calculated. It can be seen that the relationship between starting delay and distance to recovery is approximately linear. The expected value should be perfectly linear, except that $m_-$ is estimated in the loop in this simulation by running TSP under the

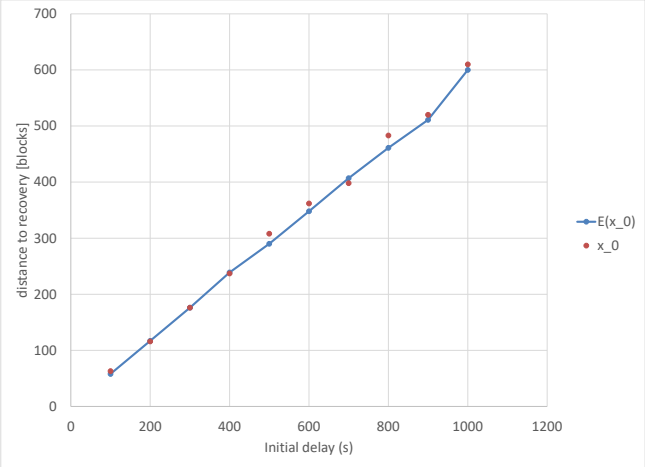



Figure 5.2: Expected and actual distance to recovery for different initial delays

same conditions.

Large disturbances were used in this example so that the results would be easier to see and interpret. Clearly recovery takes time. Most bus routes have fewer than 100 signals from end to end, so in these examples the bus would not make up the entire delay en route. Some of the delay would be made up en route, and the slack in the turnaround time would be used to make up the rest. If the turnaround time is not sufficient, the next trip would start with some initial delay and the bus would continue to recover. Passengers still benefit, as the delay decreases steadily and would be less noticeable to those passengers further removed from the disturbance.

## 5.3 Description of control algorithm

The inherent tradeoff in transfer coordination is that the out of vehicle waiting time of transferring passengers can be reduced, but at the cost of imposing delays on onboard and downstream passengers. The first condition for holding a bus is that the benefit to transferring passengers should outweigh the cost to onboard (and downstream) passengers. If this condition cannot be satisfied, then there are no good opportunities for coordination and the bus should depart as soon as it has finished boarding and alighting passengers. If there is at least one point where benefits outweigh costs within the given time horizon, coordination is feasible and the problem consists of finding the optimal holding time based on the information available. Situations with no good opportunities are really just a special case where the optimal holding time is 0, so this formulation can be used with no preconditions.

The proposed control strategy for transfer coordination is distributed, meaning that

there is no centralized controller. Each bus becomes a decision maker when it arrives at a transfer point. The decision is purely opportunistic - no attempt is made to guarantee certain transfers by coordinating multiple buses, only to take the best action for the present situation. The upside of this approach is that it can be used at every stop where there is a connecting line – and whether or not the connection is planned.

The information needed to implement this control strategy consists of estimates of upcoming bus arrivals and counts of transferring and affected passengers. The decision making bus may already have some of the real-time information it needs, like the number of passengers on board and the backward headway. The arrival times of connecting buses can be requested when the decision making bus is approaching the transfer point. Other parameters, like the uncertainty in arrival estimates, transferring passengers, and recovery, are difficult to estimate in real time so historical values can be used and stored locally.

The proposed strategy for transfer coordination is a general method which can handle complexities such as multiple connecting buses and dispersed passenger arrivals due to different walking speeds. The strategy described here was informed by the earlier development of a transfer coordination strategy for a simple case of one connecting bus and the simultaneous arrival of all transferring passengers. This simple case is introduced first and the general method comes after. The analytical formulations developed for the simple case show that uncertainty in the real-time estimates of bus arrivals does not change the optimal control strategy. Therefore, the general method assumes that all predictions made are exact. For this reason the general method is not necessarily optimal but it is likely very good.

## Simple case: two buses with synchronous transfers

The simplest case of transfer coordination is when there is one decision-making bus, one connecting bus and transferring passengers arrive all at once. The decision-maker's choice is essentially whether or not it is worthwhile to hold for the transferring passengers from this connecting bus. Despite its simplicity, this case is useful for developing analytical formulations, deriving an optimal control strategy with and without uncertainty, testing various "what-if" scenarios, and providing insights for the general case.

The notation is as follows:

### *Parameters*

**Arrival time** $A_d$ [s]: arrival time of the deciding bus at the transfer point ($A_d = 0$ from now on)
**Arrival time** $A_c$ [s]: arrival time of the connecting bus at the transfer point
**Arrival time** $\bar{A}_c$ [s]: unbiased prediction of $A_c$
**Headway** $H_d$ [s]: real time backward headway on the route of the deciding bus
**Headway** $\bar{H}_d$ [s]: unbiased prediction of $H_d$
**Standard Deviation** $\sigma_a$ [s]: standard deviation of $A_c|\bar{A}_c$ (measures the quality of the prediction)

**Standard Deviation** $\sigma_H$ [s]: standard deviation of $H_d|\bar{H}_d$ (measures the quality of the prediction)
**Affected Passengers** $P_a$ [persons]: passengers on the deciding bus who would be affected if it holds for a transfer[1]
**Affected Passengers** $\bar{P}_a$ [persons]: unbiased prediction of $P_a$
**Transferring Passengers** $P_t$ [persons]: passengers wishing to transfer to the deciding bus at the transfer point[1]
**Transferring Passengers** $\bar{P}_t$ [persons]: unbiased prediction of $P_t$
**Recovery** $\rho$ []: reflects the portion of holding time experienced by the average affected passenger on the deciding bus. $\rho = 1$ means no holding time is recovered en route, $\rho < 1$ means (1-$\rho$) of the holding time is recovered before affected passengers get off the bus. The product $\rho P_a$ is the equivalent number of passengers affected.

### *Decision Variable*

**Maximum Hold Time** $a_{max}$ [s]: maximum time that a bus should be willing to hold to receive transferring passengers; e.g. hold until $A_c$ if $\bar{A}_c \leq a_{max}$. The maximum hold time is strictly less than the backward headway: $a_{max} < H_d$.

### *Objective Function*

**Cost (a.k.a. transfer delay)** $K$ [person-s]: Cost of the chosen control action in passenger delay

## Certain Information

Let us start with a deterministic case where all real time estimates are equal to the true value. Each instance of this control strategy begins when the decision-making bus arrives at the transfer point and consults the real-time information. This occurs at time $A_d = 0$. It was mentioned above that the decision-making bus should calculate the maximum time that it is willing to hold, $a_{max}$, and then compare the estimated arrival time of the connecting bus to this value. This simple rule forms the basis of the proposed control strategy. The cost ($K$) of the control strategy is the delay to all affected passengers. The formula for cost in terms of the control variable, $a_{max}$, is as follows:

$$K(A_c, \bar{A}_c, P_a, P_t, H_d, a_{max}) = \begin{cases} \rho P_a A_c & \bar{A}_c \leq a_{max} \\ P_t(H_d - A_c) & \bar{A}_c > a_{max}. \end{cases} \tag{5.5a}$$

In the deterministic case, $A_c \equiv \bar{A}_c$, $H_d \equiv \bar{H}_d$, $P_a \equiv \bar{P}_a$, and $P_t \equiv \bar{P}_t$. Then, the variables $A_c$ and $H_d$ on the RHS of (5.5a) are known so the conditional expectation of cost is given

[1]The standard deviation is not specified because distribution and variance turn out not to matter.

by:

$$E(K|\bar{A}_c, \bar{P}_a, \bar{P}_t, \bar{H}_d, a_{max}) = \begin{cases} \rho \bar{P}_a \bar{A}_c & \bar{A}_c \leq a_{max} \\ \bar{P}_t(\bar{H}_d - \bar{A}_c) & \bar{A}_c > a_{max}. \end{cases} \tag{5.5b}$$

Of interest is the minimum of (5.5b) across all possible values of the control variable $a_{max}$. It is easy to find because the cost of holding for a transfer, $\rho \bar{P}_a \bar{A}_c$, and the cost of not holding, $\bar{P}_t(\bar{H}_d - \bar{A}_c)$, vary linearly with $\bar{A}_c$. Since we want to achieve the least of these two costs choose $a_{max}$ as the value of $\bar{A}_c$ that renders the two costs equal, i.e.:

$$a^*_{max} = \frac{\bar{P}_t \bar{H}_d}{\rho \bar{P}_a + \bar{P}_t}. \tag{5.6a}$$

Combining (5.6a) and (5.5b) yields the conditional expectation of optimal cost:

$$E(K^*|\bar{A}_c, \bar{P}_a, \bar{P}_t, \bar{H}_d) = \frac{\bar{H}_d \bar{P}_t \bar{P}_a \rho}{2(\bar{P}_t + \bar{P}_a \rho)} = \frac{1}{2} \bar{H}_d \frac{\bar{P}_t}{\bar{P}_t + \bar{P}_a \rho} \bar{P}_a \rho. \tag{5.6b}$$

The optimal control strategy is therefore to hold if $\bar{A}_c \leq a^*_{max}$. It can be seen from (5.6a) that $a^*_{max}$ is simply the backward headway on the deciding bus' route multiplied by the ratio of transferring passengers to the sum of transferring and affected passengers. (The affected passengers on the deciding bus are adjusted by the recovery factor because it is assumed that passengers only experience the schedule deviation when they get off the bus.) Therefore, if most of the holding time can be made up before the on-board passengers get off, so that $\rho$ is small then $a^*_{max}$ is large; i.e. holding becomes very attractive because $a^*_{max} \to \bar{H}_d$.

## Uncertain Information

Real bus systems, however, experience perturbations from traffic signals, mixed traffic, variable dwell times, and so on. As a result, the actual values of $A_c$, $H_d$, $P_t$, and $P_a$ may differ from the estimates $\bar{A}_c$, $\bar{H}_d$, $\bar{P}_t$, and $\bar{P}_a$ available at the time of decision. The formulas for conditional expected cost and maximum hold time have been rederived on account of the uncertainty in these estimates. We assumed the following distributions:

$$A_c|\bar{A}_c \sim U(\bar{A}_c - \sigma_a\sqrt{3}, \bar{A}_c + \sigma_a\sqrt{3})$$
$$\left.\begin{matrix} P_a|\bar{P}_a \\ P_t|\bar{P}_t \end{matrix}\right\} \text{(distribution and variance turn out not to matter)}$$
$$H_d|\bar{H}_d \sim U(\bar{H}_d - \sigma_H\sqrt{3}, \bar{H}_d + \sigma_H\sqrt{3}).$$

Uniform distributions of bus arrivals are used for simplicity. Real distributions of bus arrivals can be expected to have a strict lower bound and a long tail, e.g. a type II extreme value distribution. The reason is that buses are unlikely to be much earlier than a near-term arrival estimate but could be late due to some unexpected delay.

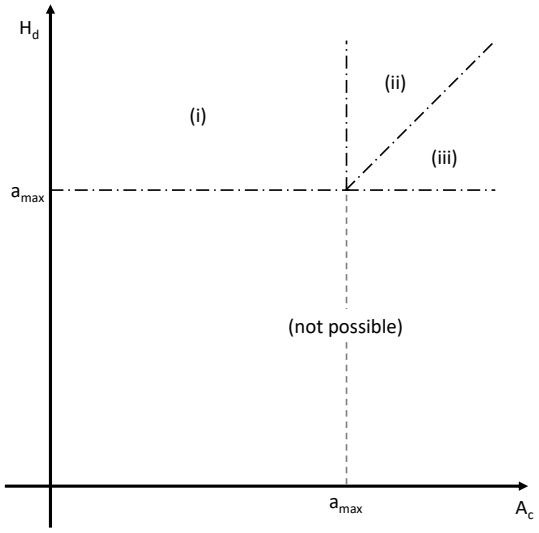



Figure 5.3: Possible outcomes with uncertainty

Two additional assumptions were made to further reduce the complexity of the analytical expressions:

$$\sigma_a\sqrt{3} \leq H_d - a^*_{max}, \tag{5.7a}$$

$$\bar{A}_c \sim U(0, \bar{H}_d). \tag{5.7b}$$

The implication of (5.7a) is that a connecting bus with which we decide to coordinate (because $\bar{A}_c \leq a^*_{max}$) is assured to arrive before $H_d$. Since $H_d$ is not known in real time, (5.7a) cannot be verified at the time of decision. The sufficient but not necessary alternative $\sigma_a\sqrt{3} \leq \bar{H}_d - \sigma_H\sqrt{3} - a^*_{max}$ could be used for this purpose. Equation (5.7b) indicates the range of estimated bus arrivals over which performance is evaluated - the chosen range reflects a scenario without preplanned coordination.

**Alternative to** (5.7a)

It is possible to relax (5.7a) by truncating transfer delay. In the deterministic case, the decision rule is to hold when $\bar{A}_c \leq a_{max}$, where $a_{max} \in [0, H_d]$ and $H_d \geq a_{max}$. Equation (5.7a) is one way of translating the conditions on $a_{max}$ to the case with uncertainty. It guarantees that $A_c \geq H_d$ by requiring that the entire distribution of $A_c$ be less than or equal to $H_d$. Particularly when $\sigma_a$ is large, (5.7a) excludes many cases where the realized value of $A_c$ turns out to be less than or equal to $H_d$.

Figure 5.3 shows the different outcomes of $H_d$ and $A_c$ with respect to $a_{max}$ that are possible with uncertain information. The region where $H_d < a_{max}$ is marked "not possible" because this outcome violates the constraints on $a_{max}$. In case (i), the deciding bus holds for

the connecting bus because $A_c \leq a_{max}$. Some of this region would be excluded by (5.7a). In cases (ii) and (iii), the deciding bus does not hold because $A_c > a_{max}$. Case (ii) represents the outcome where the connecting bus arrives before the next bus. The delay to transferring passengers is $H_d - A_c > 0$, and this cost is the responsibility of the deciding bus. Case (iii) represents the outcome where the next bus arrives before the connecting bus ($H_d < A_c$). In this case, the delay to these transferring passengers should be associated with the next bus. Then, to avoid double counting, the cost to our bus should be 0. The cost of not holding in cases (ii) and (iii) can be written $[H_d - A_c]^+$. The expression will be positive in case (ii) and negative, truncated to 0 in case (iii).

The cost function in this alternative formulation is:

$$K(A_c, P_a, P_t, H_d, a_{max}) = \begin{cases} \rho P_a A_c & if\ A_c \leq a_{max} \\ [H_d - A_c]^+ & if\ A_c > a_{max} \end{cases} \tag{5.8}$$

### All Information

The analytical formulation for conditional expected cost with uncertain information is obtained by introducing uncertainty one parameter at a time. This derivation uses the constraint in (5.7a). This process begins with Equation (5.5a):
$E(K|A_c, \bar{A}_c, P_a, P_t, H_d, \bar{H}_d, a_{max}) =$

$$\begin{cases} \rho P_a A_c & \bar{A}_c \leq a_{max} \\ P_t(H_d - A_c) & \bar{A}_c > a_{max}. \end{cases}$$

The two conditional expressions are linear functions of the parameters $P_a$ and $P_t$, respectively. As a result, the true values of these parameters can simply be replaced with their estimates, provided the estimates are unbiased.

### Everything but $A_c$

Now the aim is to write expected cost without knowing $A_c$. This can be accomplished by integrating out $A_c$ in both conditional expressions. The interval $(a_{max}, H_d]$ is subdivided into cases where the entire distribution of $A_c$ falls before $H_d$ and cases where $A_c$ can fall before or after $H_d$. The integration is as follows:
$E(K|\bar{A}_c, \sigma_a, \bar{P}_a, \bar{P}_t, H_d, \bar{H}_d, a_{max}) =$

$$\begin{cases} \int_{\bar{A}_c - \sigma_a\sqrt{3}}^{\bar{A}_c + \sigma_a\sqrt{3}} \rho \bar{P}_a A_c (\frac{1}{\sigma_a\sqrt{12}})\ dA_c & \bar{A}_c \leq a_{max} \\ \int_{\bar{A}_c - \sigma_a\sqrt{3}}^{\bar{A}_c + \sigma_a\sqrt{3}} \bar{P}_t (H_d - A_c)(\frac{1}{\sigma_a\sqrt{12}})\ dA_c & a_{max} < \bar{A}_c < H_d - \sigma_a\sqrt{3} \\ \int_{\bar{A}_c - \sigma_a\sqrt{3}}^{\bar{A}_c + \sigma_a\sqrt{3}} \bar{P}_t (\frac{4}{3}H_d - A_c)(\frac{1}{\sigma_a\sqrt{12}})\ dA_c & a_{max} < H_d - \sigma_a\sqrt{3} \leq \bar{A}_c \leq H_d, \end{cases}$$

and results in the following expected cost:

$$E(K|\bar{A}_c, \sigma_a, \bar{P}_a, \bar{P}_t, H_d, \bar{H}_d, a_{max}) =$$

$$\begin{cases} \rho\bar{P}_a\bar{A}_c & \bar{A}_c \leq a_{max} \\ \bar{P}_t(H_d - \bar{A}_c) & a_{max} < \bar{A}_c < H_d - \sigma_a\sqrt{3} \\ \bar{P}_t(\frac{4}{3}H_d - \bar{A}_c) & a_{max} < H_d - \sigma_a\sqrt{3} \leq \bar{A}_c \leq H_d. \end{cases}$$

**Everything but $A_c, \bar{A}_c$**

The next step is to eliminate $\bar{A}_c$. It is assumed that the two routes are completely uncoordinated, and that the connecting bus could arrive at any time within one headway. Now $\bar{A}_c$ can be integrated out:

$$\begin{aligned} E(K|\sigma_a, \bar{P}_a, \bar{P}_t, H_d, \bar{H}_d, a_{max}) =& \frac{1}{\bar{H}_d}\Bigg( \int_0^{a_{max}} (\rho\bar{P}_a\bar{A}_c)d\bar{A}_c + \int_{a_{max}}^{H_d-\sigma_a\sqrt{3}} (-\bar{A}_c\bar{P}_t + H_d\bar{P}_t)d\bar{A}_c \\ &+ \int_{H_d-\sigma_a\sqrt{3}}^{H_d} (-\bar{A}_c\bar{P}_t + \frac{4}{3}H_d\bar{P}_t)d\bar{A}_c \Bigg), \end{aligned}$$

to obtain the following expected cost:

$$\begin{aligned} E(K|\sigma_a, \bar{P}_a, \bar{P}_t, H_d, \bar{H}_d, a_{max}) =& \frac{1}{\bar{H}_d}\Bigg( \frac{1}{2}a_{max}^2\bar{P}_t - a_{max}H_d\bar{P}_t + \frac{1}{2}H_d^2\bar{P}_t + \frac{1}{2}a_{max}^2\bar{P}_a\rho + \frac{4}{\sqrt{3}}H_d\bar{P}_t\sigma_a \\ &- \sqrt{3}H_d\bar{P}_t\sigma_a \Bigg). \end{aligned}$$

**Uncertain everything**

Now it is possible to eliminate $H_d$, leaving only parameters that will be known at the time of decision. $H_d$ is assumed to follow a uniform distribution. Integrating over the probability density of $H_d$:

$$\begin{aligned} E(K|\sigma_a, \bar{P}_a, \bar{P}_t, \bar{H}_d, \sigma_H, a_{max}) =& \int_{\bar{H}_d-\sigma_H\sqrt{3}}^{\bar{H}_d+\sigma_H\sqrt{3}} \frac{1}{\bar{H}_d}\Bigg( \frac{1}{2}a_{max}^2\bar{P}_t - a_{max}H_d\bar{P}_t + \frac{1}{2}H_d^2\bar{P}_t + \frac{1}{2}a_{max}^2\bar{P}_a\rho \\ &+ \frac{4}{\sqrt{3}}H_d\bar{P}_t\sigma_a - \sqrt{3}H_d\bar{P}_t\sigma_a \Bigg)\left(\frac{1}{\sigma_H\sqrt{12}}\right)dH_d, \end{aligned}$$

the conditional expectation of cost with uncertainty in all parameters is as follows:

$$\begin{aligned} E(K|\sigma_a, \bar{P}_a, \bar{P}_t, \bar{H}_d, \sigma_H, a_{max}) =& -a_{max}\bar{P}_t + \frac{a_{max}^2\bar{P}_t}{2\bar{H}_d} + \frac{\bar{H}_d\bar{P}_t}{2} + \frac{a_{max}^2\bar{P}_a\rho}{2\bar{H}_d} \\ &+ \frac{\bar{P}_t\sigma_a}{\sqrt{3}} + \frac{\bar{P}_t\sigma_H^2}{2\bar{H}_d}. \end{aligned} \tag{5.9}$$

Similar to the deterministic case, the optimal control strategy is the one which minimizes (5.9) with respect to $a_{max}$. Since (5.9) is convex in $a_{max}$, the optimal $a_{max}$ is obtained by taking the derivative of (5.9) with respect to $a_{max}$ and setting it equal to zero. First, take a derivative with respect to $a_{max}$:

$$\frac{d}{da_{max}} E(K|\sigma_a, \bar{P}_a, \bar{P}_t, \bar{H}_d, \sigma_H, a_{max}) = -\bar{P}_t + \frac{a_{max}\bar{P}_t}{\bar{H}_d} + \frac{a_{max}\bar{P}_a\rho}{\bar{H}_d}.$$

By setting the derivative equal to 0 and solving for $a_{max}$, an expression is obtained for the maximum holding time in terms of real-time information:

$$a^*_{max} = \frac{\bar{P}_t\bar{H}_d}{\rho\bar{P}_a + \bar{P}_t}, \tag{5.11a}$$

which is substituted into (5.9) to obtain the conditional expectation of optimal cost with uncertain information:

$$E(K^*|\sigma_a, \bar{P}_a, \bar{P}_t, \bar{H}_d, \sigma_H) = \frac{\bar{P}_t(3\bar{H}_d^2\bar{P}_a\rho + 2\sqrt{3}\bar{H}_d(\bar{P}_t + \bar{P}_a\rho)\sigma_a + 3(\bar{P}_t + \bar{P}_a\rho)\sigma_H^2)}{6\bar{H}_d(\bar{P}_t + \bar{P}_a\rho)}. \tag{5.11b}$$

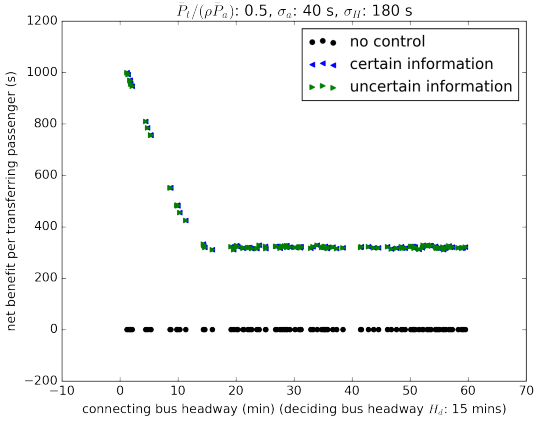
This expression can be rearranged:

$$\begin{aligned} E(K^*|\sigma_a, \bar{P}_a, \bar{P}_t, \bar{H}_d, \sigma_H) =& \frac{1}{6\bar{H}_d(\bar{P}_t + \bar{P}_a\rho)}\Big(\bar{P}_t(3\bar{H}_d^2\bar{P}_a\rho \\ &+ 2\sqrt{3}\bar{H}_d(\bar{P}_t + \bar{P}_a\rho)\sigma_a \\ &+ 3(\bar{P}_t + \bar{P}_a\rho)\sigma_H^2)\Big) \end{aligned} \tag{5.12}$$

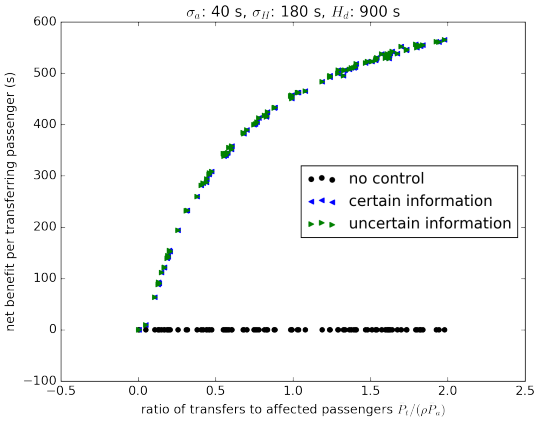
The first line represents the deterministic cost, while the second line represents the added cost due to uncertainty in the connecting bus arrival, and the third line the added cost due to uncertainty in the headway.

We note that the expression in (5.11a) is identical to (5.6a). This fact is very useful. It says that in making a control decision *uncertainty can be ignored* provided we use as inputs unbiased estimates of the various parameters. The simulation results in Figure 5.4 confirm this fact. They also match the analytic formulations, confirming the soundness of the derivations leading to (5.6a) and (5.11a). Most importantly, this finding means that the deterministic formulation can probably be used with no loss of performance in more complex scenarios. The deterministic formulation has fewer parameters to estimate and is generally easier to implement.

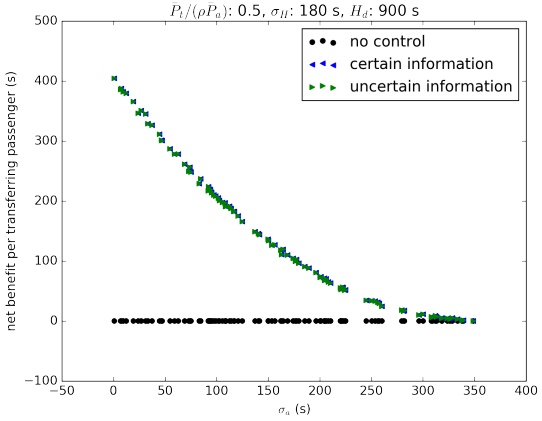
The figure also illustrates the benefit of the control. It compares deterministic and uncertainty control under various parameter combinations. Each point in these scatter plots is the average of 10,000 draws. One parameter varies per plot, the other parameters are held constant with the values given in the plot title. Projected and actual arrival times of the connecting bus, $\bar{A}_c$ and $A_c$, are random variables drawn from the specified distributions.



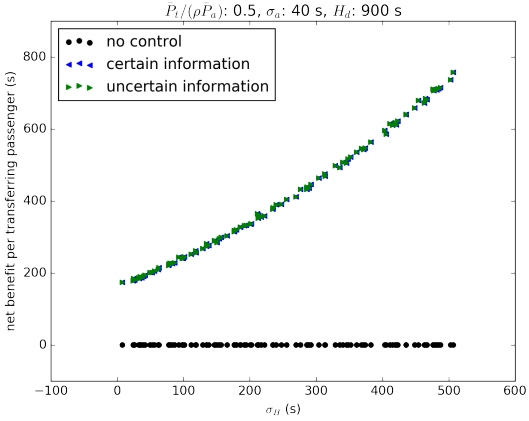

(a) Simulations with different connecting route headway (note: deterministic and uncertainty controls overlap)




(b) Simulations varying the ratio of transferring to affected passengers




(c) Simulations with different values of $\sigma_a$




(d) Simulations with different values of $\sigma_H$

Figure 5.4: Comparison of the control with and without uncertainty

## General case

The preceding section described a simple case of transfer coordination where coordination is done with only one connecting bus. A general method will now be developed which can handle additional complexities such as multiple connecting buses and dispersed passenger arrivals due to different walking speeds. The general method uses the finding that uncertainty does not change the optimal control strategy in the simple case. For this reason the general method is not necessarily optimal but it is likely very good. In essence, it will be assumed that all the predictions made are exact. For this reason, in this section all the overbars (used to distinguish estimates from realized values) are eliminated. The notation is the same as the simple case with two additions. Subscripts are now needed on $A_c$ and $P_t$ because there can be multiple connecting buses, and there is one new term to define:

**Passenger Arrival Curve** $N(t)$ [persons]: cumulative arrivals of transferring passengers to time $t$

To implement the generalization, let us introduce the notion of a passenger arrival curve. Each instance of this control strategy begins when the decision-making bus arrives at the transfer point and consults the real-time information. This occurs at time $A_d = 0$, where $A_d$ represents the arrival time of the deciding bus. The real-time information on connecting trips is used to construct passenger arrival curves. The arrival curve for a connecting bus $i$, denoted $N_i(t)$, is obtained by taking the most likely arrival time, $A_{ci}$ and drawing the cumulative distribution function of passenger arrivals, scaled by the expected number of transfers $P_{ti}$. The cumulative distribution function of passenger arrivals could be a step function if it is assumed that all passengers arrive at once, or it could reflect the distribution of passenger walking times. Finally, the arrival curves of all connecting buses are added to obtain the total arrival curve for connecting passengers, $N(t) = \sum_i N_i(t)$, which is used by the decision-making bus to determine its holding time. An example is shown in Figure 5.5. In the figure, there are three connecting trips with different numbers of transferring passengers. Passenger walking times are uniformly distributed between 30 and 150 s for all three buses.

The shaded area in Figure 5.5 shows the delay to transferring passengers if the bus is held until $t = A_c$, where $A_c$ is a control variable representing the holding time. Therefore the delay saved to transferring passengers is the area of the white rectangle, $N(A_c) \cdot (H_d - A_c)$, where $H_d$ is the real-time headway the decision-making bus and the next bus on the route. The delay to affected passengers is $\rho P_a A_c$, where $\rho$ is the recovery factor and $P_a$ is the number of affected passengers (who could be on board or downstream). The $\rho$ parameter represents the fraction of holding delay remaining when the average affected passenger gets off the bus. It is assumed that passengers are indifferent to holding time that is made up en route. Thus, the objective function to be minimized is:

$$K(A_c) = \rho P_a A_c - N(A_c)(H_d - A_c). \tag{5.13}$$

This function can be minimized over $A_c$ to obtain the optimal holding time, $A_c^*$.

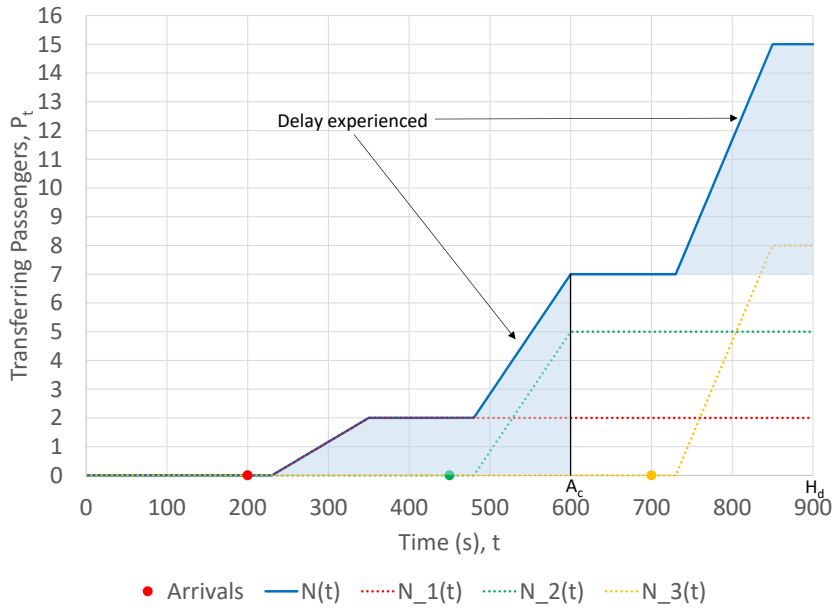
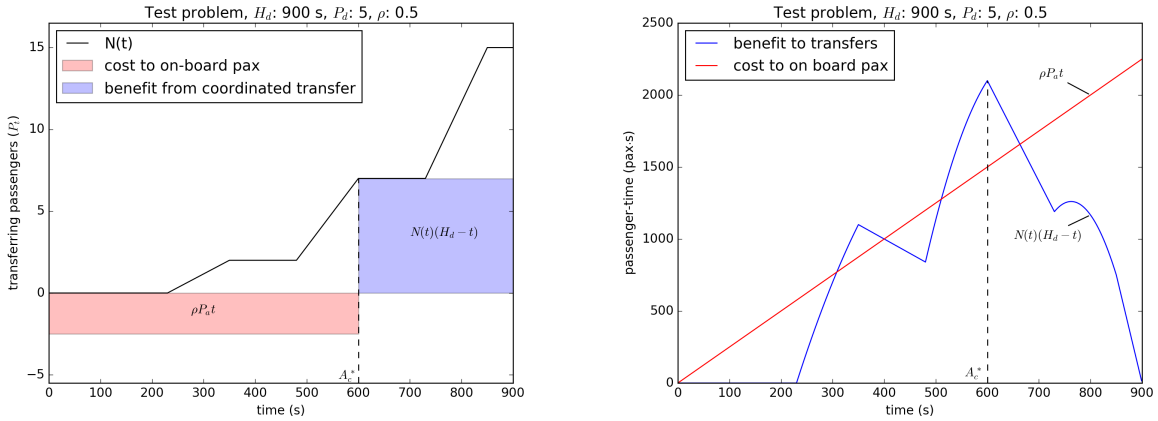



Figure 5.5: Arrival curve construction and delay if holding until $A_c$

The shaded rectangles in Figure 5.6(a) show the optimal solution in the example of Figure 5.5, and Figure 5.6(b) confirms that the control rule minimizes transfer delay, as the benefit to transferring passengers of holding for 600 s exceeds the cost to affected passengers by the most of any point within the time horizon. Note, it is possible for $A_c^* = 0$ if $\rho P_a$ is so large that $\rho P_a t > N(t)(H_d - t), \forall t$.

## Downstream Effects

The objective function in (5.13) is easy to work with. However, much like the number of transferring passengers depends on the holding time, so does the number of affected passengers. It is possible to replace the $\rho P_a A_c$ term with an affected passenger curve. An advantage of doing this is that it replaces the $\rho$ parameter, which combines the rate of recovery and the trip length of onboard passengers, with $m_-$, which is the rate of recovery and is much easier to observe. Affected passengers include passengers already onboard at the transfer point and those waiting at stops downstream. Both will experience some delay less than or equal to $A_c$, depending on how far downstream they alight or board, respectively. When the bus uses CSP to recover its holding delay, the cost of holding to all affected




(a) Arrival curve showing optimal solution (b) Plot of cost/benefit of transfer coordination

Figure 5.6: Graphical Explanation of Arrival Curve Method

passengers is given by the summation:

$$K_a(A_c) = \sum_{i=1}^{\lfloor A_c/m_- \rfloor} (A_c - i \cdot m_-)P_i, \tag{5.14}$$

where $P_i$ represents the number of passengers boarding and alighting at the $i$th station. Passenger arrivals at bus stations are often represented by the Poisson distribution, so let us say that $P_i \sim \text{Pois}(\lambda_i)$. If the downstream segment of the route is relatively homogeneous, then it is reasonable to assume that $\lambda_1 = \lambda_2 = ... = \lambda_{\lfloor A_c/m_- \rfloor}$. Using this assumption, the expected value of $K_a(A_c)$ can be written:

$$E[K_a(A_c)] = \frac{1}{2}(A_c - m_-)\left(\frac{A_c}{m_-}\right)\lambda. \tag{5.15}$$

The objective function with endogenous $P_a$ is:

$$K_a(A_c) = \left(\frac{A_c^2}{2m_-} - \frac{A_c}{2}\right)\lambda - N(A_c)(H_d - A_c). \tag{5.16}$$

As before, this function can be minimized over $A_c$ to obtain the optimal holding time, $A_c^*$.

## 5.4 Evaluation

### Case Study

In order to test the proposed control scheme in a real-world scenario, we collected data at a subway-to-bus transfer point in Oakland, California. The Rockridge BART station is

elevated and the main exit leads directly to the westbound bus stop. This bus stop is the starting point for AC Transit's 51B route, which heads west to the UC Berkeley campus and west Berkeley. Two observers were used, one at the exit to the BART station and one at the bus stop. The observer at the exit started shooting video when a train arrived (which could be observed because the tracks are elevated) and stopped when all the passengers had exited the station. The observer at the bus stop counted transferring passengers and shot video of the bus boarding process. At the same time, a script was running that requested information on upcoming trips from the BART and AC Transit APIs once a minute. This experiment obtained the time that every BART passenger exited the station, the numbers of transferring and total passengers boarding each bus, and forecasted arrival times for each BART and bus trip over time.

These data were used to associate BART passengers to trains and determine the walking time from train platform to bus stop. The station exit and bus boarding videos were compared to confirm the number of transferring passengers and determine their arrival time at the bus stop. A histogram of train platform to bus stop walking times is shown in Figure 5.7. The mean walking time is 93 s, with a standard deviation of 43 s. In 160 observations, 82% fall between 30-120 s, and 95% of the observations are contained between 30 s and 150 s.

The API data was used to evaluate the quality of the estimated arrival times and to estimate the parameters $\sigma_a$ and $\sigma_H$. Although these parameters are not used in the general method, they allow for a comparison of the uncertainty in bus/train arrival estimates and passenger walking time. The standard deviation of bus arrival estimates depends on the prediction horizon. AC Transit's real-time estimates begin around 30 minutes before arrival and the standard deviation of all estimates is $\sigma_H = 66$ s. The control strategy is only concerned with estimates within one headway, 10 minutes in this case, and this subset has $\sigma_H = 31$ s. Real-time estimates for BART only begin 15-20 minutes before arrival and have a standard deviation of $\sigma_a = 30$ s. The standard deviation of walking times (43 s) is greater than both BART and AC Transit (within 10 mins), suggesting that the dispersion of passenger arrivals due to differences in walking time is important and should not be ignored.

The data collected were then used to virtually apply the proposed control strategy. The estimated arrival times, $\bar{A}_c$ and $\bar{H}_d$, are directly observed in the API data (overbars are used in the case study to distinguish estimates from realized values). Because this stop is the beginning of the bus route, there are no through passengers on board. Instead, passengers who were waiting at the stop when the bus arrived were counted as the affected passengers, $P_a$. The remaining unknown is the number of transferring passengers, $\bar{P}_t$. In the data, each BART train contains between 0-4 transferring passengers, with 9 out of 12 trains containing at least one transferring passenger. We used the average, $\bar{P}_t = 2$.

Figure 5.8 shows the cost/benefit plots with the estimates that the bus would have at the time of decision. Clearly the optimal action for the buses at 8:14, 8:33, and 8:45 is to depart immediately since the origin is the only point where the benefit to transfers, $N(t)(H_d - t)$, is greater than or equal to the cost to affected passengers, $\rho P_a t$. The optimal action for the bus at 8:21 is to hold for the first group of arriving passengers and then depart.

Table 5.1 shows the real time estimates and the observed values. The first two columns

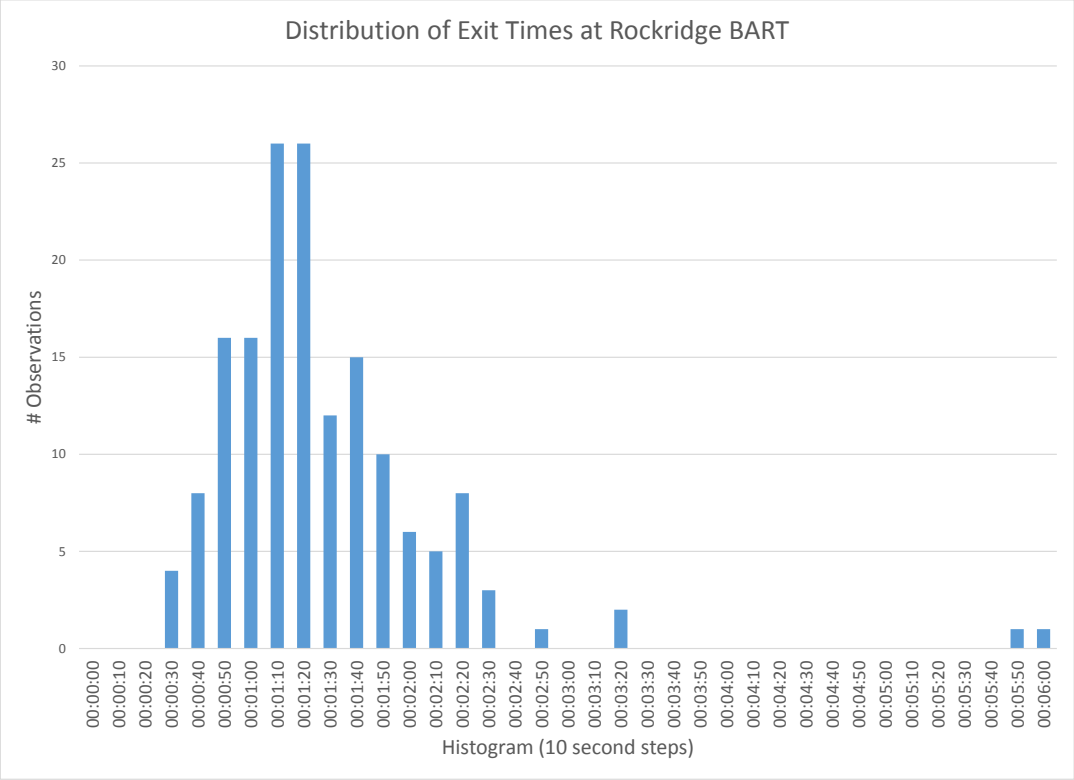
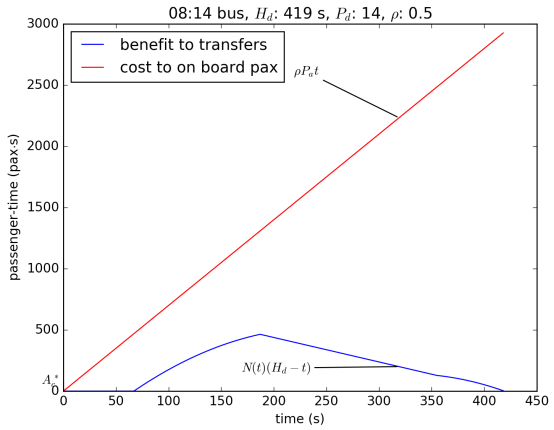


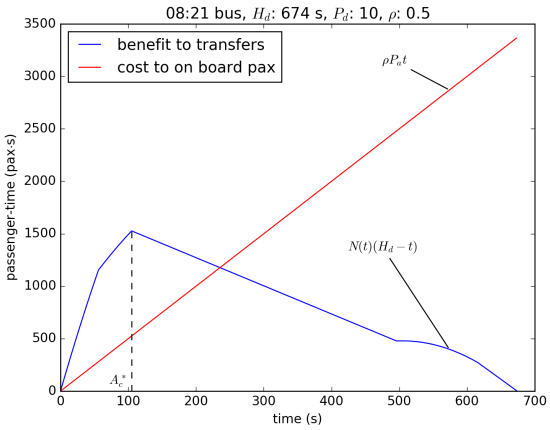

Figure 5.7: Histogram of walking times

show the information about the deciding bus. The information about the connecting trains is divided into two groups: the real-time estimates available to the deciding bus and the observed values. The positive $\bar{A}_c$ values are all multiples of 60 s because the precision of the API arrival estimates is in minutes. The control action column is the action obtained from applying the general method with the real-time estimates (as seen in Figure 5.8), while the perfect action column is the action obtained from the true values. In this dataset, they are the same. The benefits of holding the 8:21 bus for the SF Airport train would actually have been greater than expected because it had 3 transferring passengers instead of the 2 predicted.

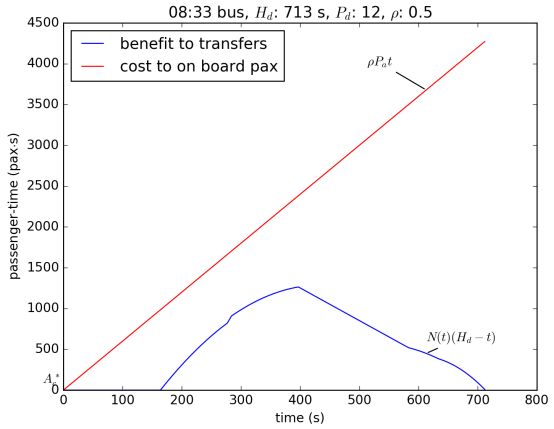
The two negative $\bar{A}_c$ values are an interesting case where the connecting trains arrived before the bus but the transferring passengers were still en route due to the size of the station. The Pittsburg/Bay Point train is listed for both the 8:14 and 8:21 buses because the passenger arrival curve predicted that one passenger would arrive before the 8:21 bus and



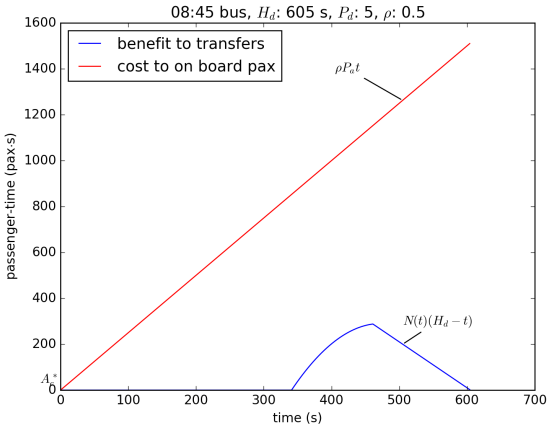

(a) Cost/benefit plot for 8:14 bus




(b) Cost/benefit plot for 8:21 bus




(c) Cost/benefit plot for 8:33 bus




(d) Cost/benefit plot for 8:45 bus

Figure 5.8: Arrival curve representations of the four scenarios

the other after. The prediction was correct, one passenger got on the 8:21 bus and the other had to wait for the 8:33 bus. Using only the mean walk time would have predicted that both passengers got on the 8:21 bus without holding. This example shows why the general method is needed. In large stations it is possible for walking time to be the difference between making and missing a transfer. Only by including it in the control strategy can these situations be modeled correctly.

Table 5.2 compares the cost of the no control case with the general method for two different values of $\rho$. It is apparent that the 8:21 bus is the best candidate for holding because the cost to transferring passengers of waiting for the next bus in the no control case is the largest by far. Holding the 8:21 bus for the SF Airport train reduces transfer delay (inclusive of the costs of holding) by 30% if the holding time is not made up ($\rho = 1.0$) and by 39% if half the holding time can be made up before the mean affected passenger gets off

Table 5.1: Action plan

| Deciding bus | | | Real-time estimates | | | Actual | | |
|---|---|---|---|---|---|---|---|---|
| time | $P_a$ | train | $\bar{A}_c$ | $\bar{P}_t$ | control action | $A_c$ | $P_t$ | perfect action |
| 08:14 | 14 | Daly City | 60 | 2 | depart | 37 | 2 | depart |
| | | Pittsburg/Bay Point | 420 | 1 | depart | 325 | 1 | depart |
| 08:21 | 10 | Pittsburg/Bay Point | -94 | 1 | hold | -94 | 1 | hold |
| | | SF Airport | -45 | 2 | hold | -45 | 3 | hold |
| | | Daly City | 540 | 2 | depart | 466 | 2 | depart |
| 08:33 | 12 | SF Airport | 180 | 2 | depart | 134 | 2 | depart |
| | | Pittsburg/Bay Point | 240 | 2 | depart | 247 | 0 | depart |
| | | Daly City | 720 | 2 | depart | 553 | 1 | depart |
| 08:45 | 5 | SF Airport | 360 | 2 | depart | 311 | 1 | depart |
| 08:55 | 7 | | | | | | | |

Table 5.2: Out of vehicle delay (min)

| | no control | | control ($\rho = 1.0$) | | control ($\rho = 0.5$) | |
|---|---|---|---|---|---|---|
| | affected | transfer | affected | transfer | affected | transfer |
| 8:14 bus | 0.0 | 10.9 | 0.0 | 10.9 | 0.0 | 10.9 |
| 8:21 bus | 0.0 | 49.4 | 14.5 | 10.2 | 7.2 | 10.2 |
| 8:33 bus | 0.0 | 17.8 | 0.0 | 17.8 | 0.0 | 17.8 |
| 8:45 bus | 0.0 | 3.9 | 0.0 | 3.9 | 0.0 | 3.9 |
| Total | | 82.0 | | 57.3 | | 50.1 |
| Savings | | | | 30% | | 39% |

($\rho = 0.5$).

## Simulation

The case study showed that the general method can significantly reduce transfer delay in a real-world scenario. To gain more general insights about the expected benefits of transfer coordination, it is necessary to test a wider range of parameter values. As mentioned in the previous section, analytical formulations were developed for a simple case with one connecting bus and the simultaneous arrival of all transferring passengers. The expressions for expected cost in this simple case can also be used to obtain insights about the general case. The general method targets a specific point on the passenger arrival curve, $(A_c^*, N(A_c^*))$ for coordination, which is roughly equivalent to holding for a connecting bus expected to arrive at $\bar{A}_c$ with $\bar{P}_t$ transferring passengers in the simple case. Notably, these expressions reflect uncertain information because uncertainty affects cost even though it does not change the optimal control strategy. Overbars represent mean values and $\sigma_a$ and $\sigma_H$ represent the standard deviations in the distributions of the arrival time of the connecting bus and the

real-time headway, respectively.

The two expressions used are the conditional expectation of optimal cost with uncertain information, denoted $K^*$,

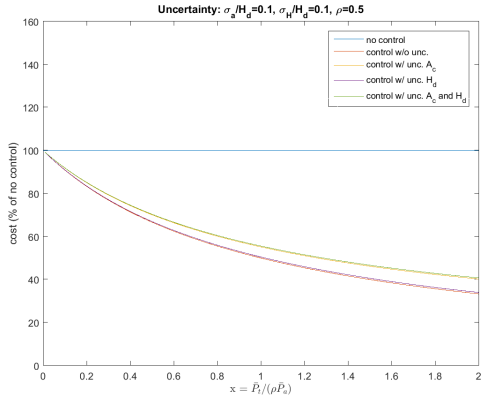
$$E(K^*|\sigma_a, \bar{P}_a, \bar{P}_t, \bar{H}_d, \sigma_H) = \frac{\bar{P}_t(3\bar{H}_d^2\bar{P}_a\rho + 2\sqrt{3}\bar{H}_d(\bar{P}_t + \bar{P}_a\rho)\sigma_a + 3(\bar{P}_t + \bar{P}_a\rho)\sigma_H^2)}{6\bar{H}_d(\bar{P}_t + \bar{P}_a\rho)}, \quad (5.17)$$

and the conditional expected cost of no control with uncertain information, denoted $K_{nc}$:

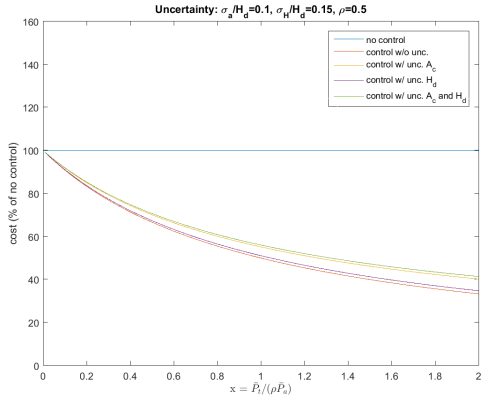
$$E(K_{nc}|\sigma_a, \bar{P}_a, \bar{P}_t, \bar{H}_d, \sigma_H) = \frac{\bar{H}_d\bar{P}_t}{2} + \frac{\bar{P}_t\sigma_a}{\sqrt{3}} + \frac{\bar{P}_t\sigma_H^2}{2\bar{H}_d}. \quad (5.18)$$

Dimensionless plots showing 5.17 and 5.18 over a range of parameter values are given in Figure 5.9. The values of $\sigma_a$ and $\sigma_H$ when considered are given in the captions. The y-axis is transfer delay as a percentage of no control, so control with uncertainty in $A_c$ alone is compared to no control with uncertainty in $A_c$ alone, and so on. The interesting result in subplots (a) and (b) is that the proposed control scheme is always better than or equal to no control, with or without uncertainty in the bus arrival estimates, and for any value of $\bar{P}_t/(\rho\bar{P}_a)$. The percent reduction in transfer delay depends primarily on this ratio, with reductions on the order of 50-60% possible if transferring passengers outnumber affected passengers. Larger values of $\sigma_a$ increase transfer delay, but do not change the overall shape of the curve. The value of $\sigma_H$ has little effect on transfer delay. Uncertainty of any kind does not undermine the control strategy as it remains superior to no control. What is lost are the potential benefits that could be realized with perfect information.

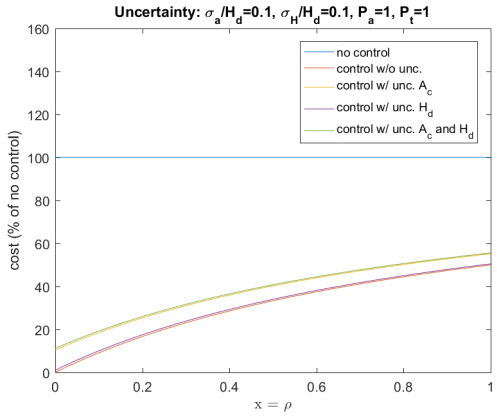
The recovery parameter $\rho$ is one of the most important as well as the mechanism by which conditional signal priority affects transfer coordination. As the rate of recovery from delays in conditional signal priority increases, $\rho$ decreases, enabling the use of transfer coordination in more cases. The effect of $\rho$ can already be seen in subplots (a) and (b), as a decrease in $\rho$ would increase the ratio $\frac{\bar{P}_t}{\rho\bar{P}_a}$, all else equal. Subplots (c) and (d) isolate the effect of $\rho$, using the same parameters as (a) and (b), respectively, with $\bar{P}_t$ and $\bar{P}_a$ fixed at unity.



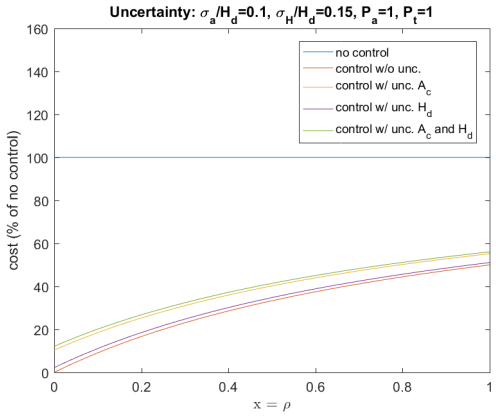

(a) Passenger plot with $\sigma_a/H_d = 0.10$ and $\sigma_H/H_d = 0.10$




(b) Passenger plot with $\sigma_a/H_d = 0.10$ and $\sigma_H/H_d = 0.15$




(c) Rho plot with $\sigma_a/H_d = 0.10$ and $\sigma_H/H_d = 0.10$




(d) Rho plot with $\sigma_a/H_d = 0.10$ and $\sigma_H/H_d = 0.15$

Figure 5.9: Equations (5.17) and (5.18) plotted over a range of parameter values

## 5.5 Macroscopic benefits

Like conditional signal priority, this method of transfer coordination is designed to run in the background at all eligible locations. Transfer coordination is a simple control strategy and most of its inputs are normally collected as part of bus operations. Some of these are monitored in real time while others could be estimated from fare receipts. It requires little computational power to check for coordination opportunities at every transfer point. An activation at a low-volume unscheduled transfer point may be unlikely, but there is no harm in considering it.

If transfer coordination were implemented systemwide in a transit agency such as San Francisco Muni, we expect that several patterns would emerge. First, most of the activations would happen at a few key locations such as downtown stops where several routes cross, large

bus stations like the Transbay Terminal, and rail stations. These locations are likely to have a high volume of transferring passengers and a connecting arrival in the next minute or two. The strategy would tend to preserve timed transfers because the bus arrivals are usually close. Other favorable situations would be those where the costs of holding are lower. Some locations have few through passengers because most people get off or transfer. An express route before a nonstop segment would work well because delay can be recovered before reaching downstream stops. Or, a point near the end of a route might be a good candidate because there are few downstream passengers. The data from transfer coordination could also reveal unexpected transfers. The transit agency might use this information to modify the bus schedule or to provide more amenities (like shelters and departure boards) at commonly used transfer points.

Second, transfer coordination would reduce the expectation and variance of transfer times by removing some of the highest values. The worst-case travel times for passengers occur when they have narrowly missed a bus. Not much can be done to solve this problem at the origin stop because the bus does not know how many passengers are on the way and when they will arrive. In transfer coordination, near misses are a good candidate because it does not take much holding time and the transferring passengers will save almost an entire headway. Changing the distribution of transfer times will make trips shorter and more reliable and could potentially affect routing decisions. If transfers are unreliable, a longer trip on high-frequency routes could be preferable to a shorter trip on low-frequency routes because the 95th percentile travel time of the former is better. With lower variance of transfer times, passengers would prefer the latter.

Third, transfer coordination can be useful in disrupted operations or during special events. Service disruptions and special events produce headways and passenger loads that are much different than normal operations. The value of running an opportunistic strategy in the background at all time is that it can intervene in these types of situations and coordinate transfers that would not normally be worthwhile. In service disruptions, transfer coordination could help to stabilize passenger travel times despite large schedule deviations. In managing special events, transfer coordination would act to avoid making a large group of people wait a long time for a transfer. Without control or extra service, long wait times are likely because many events take place on weekends or end late at night.

# Chapter 6

# Discussion

This work addresses the problem of travel time variability in bus transit systems and its consequences for passengers, namely an increase in out-of-vehicle waiting time due to late buses and missed transfers. The core of this work is a strategy of conditional signal priority, or CSP.

Conditional signal priority is a distributed control strategy that is run by individual buses at all equipped signals. It employs a simple control rule where the bus' current deviation from schedule is used to decide between two actions: requesting priority (if the bus is late) or taking no action (if the bus is on time or early). Variations with additional request types and thresholds are possible, but in general the main control guide is the bus schedule. CSP is essentially a mixture of no control and transit signal priority (TSP). Like any mixture, the range of possible outcomes is from a 0% activation rate (identical to no control) to a 100% activation rate (identical to TSP). Therefore, the schedule pace should be faster than an uncontrolled bus route and slower than a bus route using TSP. Although any schedule pace in this range is achievable, they are not all equally good. In Chapter 4, it is argued that the motion of a single CSP-controlled bus meets all of the conditions for Brownian motion with drift. Brownian motion is then used to develop formulas for the expectation and variance of schedule deviation in steady state. Simulations confirm that CSP indeed guides buses to a steady state and that the Brownian motion formulas accurately predict performance once it is reached. A comparison with no control and TSP shows that CSP has better schedule adherence both before and after it reaches steady state. No control and TSP never reach steady state. In chapter 5, Brownian motion is used to develop formulas for the expectation and variance of distance to recovery after a delay. The existence and performance of this delay recovery mechanism are confirmed in simulation.

The ability of CSP-controlled buses to recover from delays of any size allows for complementary strategies. To this end, an optimal control strategy for transfer coordination is developed. It is a distributed strategy where a bus calculates its optimal holding time by comparing the expected arrival and benefits for transferring passengers to the cost of holding for onboard and downstream passengers. This strategy advances the state of practice by considering multiple connecting buses, the impact of uncertain information, and down-

stream effects. The uncertainty in real-time information is shown not to affect the optimal decision, although it does increase cost. Simulations for a wide range of parameter values show that the transfer coordination method has a positive or zero impact on passengers but is never harmful, even with uncertain information. A case study developed from real-world observations at a transfer point in Oakland shows that the transfer coordination method could reduce transfer delay by 30% or more including the cost of holding.

The rest of this chapter contains some notes for implementation and directions for future work. The next section describes how information that is already being collected for conditional signal priority and transfer coordination can be used to benefit pedestrians. The following section describes how conditional signal priority copes with unexpected congestion on part of a bus route. The last section focuses on one direction for future work: a variant of conditional signal priority for systems where pairing is of concern.

## 6.1 Coordinating pedestrian crossings

When conditional signal priority is implemented, buses at a minimum must know their own location. For transfer coordination, these bus locations need to be communicated to other buses along with some count or estimate of passengers on board. Regardless of how it is done, information on bus locations and passenger counts is present in the system. This information could also be used to anticipate pedestrian crossing demand, which could improve safety and reduce delays for pedestrians and buses.

There are likely to be spikes in pedestrian demand at street crossings immediately before and after the arrival of a bus at a station. When the bus is approaching, there may be some pedestrians who want to cross the street to reach the station and get on the bus, as indicated in Figure 6.1. If there is no opportunity to cross the street before the bus arrives because the signal is red for these pedestrians, they may miss the bus or cross unsafely. Granting a short pedestrian phase before the bus arrives could be beneficial if it prevents someone from missing the bus or improves traffic safety.

After the bus leaves, the passengers who alighted may also want to cross the street. If the station is a transfer point and the connecting bus stop is across the street, granting another

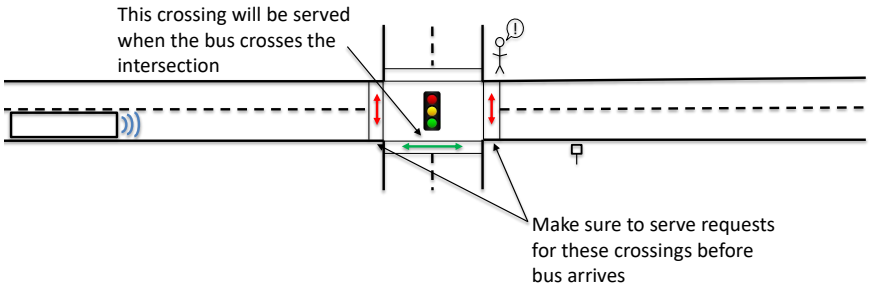



Figure 6.1: Complementary pedestrian measure on a single route

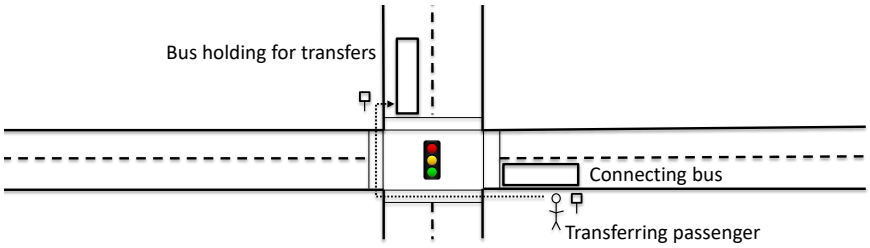



Figure 6.2: Complementary pedestrian measure at a transfer point

pedestrian phase could be instrumental in reducing holding time if there is a bus waiting. The transfer case is illustrated in Figure 6.2. The disruption to cars is easily justifiable in the first case if the pedestrian phase prevents someone from missing the bus, particularly when headways are long (larger benefits, fewer disruptions). The only difficulty may be in identifying the situations where an intervention is worthwhile, as pedestrians trying to catch a bus could be difficult to distinguish from background pedestrian demand. Intervention in the second case is justifiable if the pedestrian phase reduces holding time and schedule delay. In the US, there is a precedent of school buses, which not only drop passengers off but stop traffic to help them cross the street.

## 6.2 Unexpected congestion / other constraints

The properties of CSP are useful for reacting to unexpected congestion on a part of the bus route or for integrating it with other signal control strategies. First, it is only unexpected congestion that poses problems for buses. Bus schedules are routinely adjusted to account for recurring congestion along some or all of the route. When unexpected congestion is encountered, a bus can request priority at all signals in the congested area but it may still accumulate some delay. This delay can be made up after it leaves the congested area (if only part of the route is affected), and the time to recovery will be described by the formulas in Chapter 5.

For integration with other signal control strategies, it is useful that CSP does not expect to recieve priority all the time. It is fully expected that some requests may be denied due to signal timing constraints, conflicting bus priority requests, etc. In the worst case, priority may be unavailable at several consecutive signals due to another strategy like offset coordination that cannot be interrupted without undermining it. Even in this situation, the bus may benefit from the offset coordination. If it does not, some delay will be accumulated which can be recovered in the usual way after leaving the area of conflict. As with congestion, if conflicts with other control strategies are a regular occurrence, this can be reflected in the schedule.

## 6.3 Headway-based service

Conditional signal priority has been described in this work in a general way that can be applied to any bus route with compatible signals. This includes buses in headway-based service, although considering only one bus at a time neglects the pairing effect. CSP will react in the correct way when pairing begins to pull buses off schedule: those with short forward headways will be denied priority, and those with long forward headways will request it. Nevertheless, depending on the strength of the pairing effect, CSP may not be able to overcome it.

To address this issue, one direction for future work is the development of a variant of CSP for situations where pairing is of concern. This variant will consider forward and backward headway in its objective function in addition to schedule deviation. Buses will then be able to request priority at the next signal or hold at the next station, depending on the sign of the objective function. It is expected that this control strategy will be able to stabilize headways as well as maintain a schedule. These properties will be useful for routes with both a printed schedule and pairing effects and help to reduce operating costs in systems without a printed schedule.

# References


Abkowitz, M., Josef, R., Tozzi, J., and Driscoll, M. K. (1987). Operational feasibility of timed transfer in transit systems. *Journal of transportation engineering*, 113(2):168–177.

AhmadBeygi, S., Cohn, A., Guan, Y., and Belobaba, P. (2008). Analysis of the potential for delay propagation in passenger airline networks. *Journal of air transport management*, 14(5):221–236.

Aloulou, M. A., Haouari, M., and Mansour, F. Z. (2013). A model for enhancing robustness of aircraft and passenger connections. *Transportation Research Part C: Emerging Technologies*, 32:48–60.

Arıkan, U., Gürel, S., and Aktürk, M. S. (2016). Integrated aircraft and passenger recovery with cruise time controllability. *Annals of Operations Research*, 236(2):295–317.

Arribas-Bel, D. and Sanz-Gracia, F. (2014). The validity of the monocentric city model in a polycentric age: Us metropolitan areas in 1990, 2000 and 2010. *Urban Geography*, 35(7):980–997.

Bachelier, L. (1900). *Théorie de la spéculation*. PhD thesis, Université de Paris.

Ben-Akiva, M. E. and Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand*, volume 9. MIT press.

Berry, B. J. and Kim, H.-M. (1993). Challenges to the monocentric model. *Geographical analysis*, 25(1):1–4.

Bhat, C. R. and Sardesai, R. (2006). The impact of stop-making and travel time reliability on commute mode choice. *Transportation Research Part B: Methodological*, 40(9):709–730.

Bookbinder, J. H. and Desilets, A. (1992). Transfer optimization in a transit network. *Transportation science*, 26(2):106–118.

Bratu, S. and Barnhart, C. (2006). Flight operations recovery: New approaches considering passenger recovery. *Journal of Scheduling*, 9(3):279–298.

Brown, R. (1828). A brief account of microscopical observations made in the months of june, july and august, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Philosophical Magazine*, 4:161–173.

Cacchiani, V., Huisman, D., Kidd, M., Kroon, L., Toth, P., Veelenturf, L., and Wagenaar, J. (2014). An overview of recovery models and algorithms for real-time railway rescheduling. *Transportation Research Part B: Methodological*, 63:15–37.

Cadarso, L. and Marín, Á. (2013). Robust passenger oriented timetable and fleet assignment integration in airline planning. *Journal of Air Transport Management*, 26:44–49.

Chiu, C., Chou, C., Lee, J. H.-M., Leung, H.-f., and Leung, Y. (2002). A constraint-based interactive train rescheduling tool. *Constraints*, 7(2):167–198.

Chow, A. H. and Li, S. (2017). Modelling and managing bus service regularity with influence of prevailing traffic. *Transportmetrica B: Transport Dynamics*, pages 1–25.

Chow, A. H., Li, S., and Zhong, R. (2017). Multi-objective optimal control formulations for bus service reliability with traffic signals. *Transportation Research Part B: Methodological*, 103:248–268.

Corman, F., D'Ariano, A., Hansen, I. A., Pacciarelli, D., and Pranzo, M. (2011). Dispatching trains during seriously disrupted traffic situations. In *2011 IEEE International Conference on Networking, Sensing and Control (ICNSC)*, pages 323–328. IEEE.

Daamen, W., Lee, Y.-c., and Wiggenraad, P. (2008). Boarding and alighting experiments: Overview of setup and performance and some preliminary results. *Transportation Research Record: Journal of the Transportation Research Board*, 2042:71–81.

Daganzo, C. F. (2009). A headway-based approach to eliminate bus bunching: Systematic analysis and comparisons. *Transportation Research Part B: Methodological*, 43(10):913–921.

Daganzo, C. F. (2017). Unpublished research notes. VIA Analytics.

de Palma, A. and Picard, N. (2005). Route choice decision under travel time uncertainty. *Transportation Research Part A: Policy and Practice*, 39(4):295–324.

Delgado, F., Contreras, N., and Muoz, J. C. (2013). Holding for transfers. In *92nd Annual Transportation Research Board Conference*, Washington, D.C.

Dessouky, M., Hall, R., Zhang, L., and Singh, A. (2003). Real-time control of buses for schedule coordination at a terminal. *Transportation Research Part A: Policy and Practice*, 37(2):145–164.

Dollevoet, T., Corman, F., DAriano, A., and Huisman, D. (2014). An iterative optimization framework for delay management and train scheduling. *Flexible Services and Manufacturing Journal*, 26(4):490–515.

Dube, L., Schmitt, B. H., and Leclerc, F. (1991). Consumers' affective response to delays at different phases of a service delivery1. *Journal of Applied Social Psychology*, 21(10):810–820.

Duerr, P. (2000). Dynamic right-of-way for transit vehicles: integrated modeling approach for optimizing signal control on mixed traffic arterials. *Transportation Research Record: Journal of the Transportation Research Board*, 1731:31–39.

Dziekan, K. and Kottenhoff, K. (2007). Dynamic at-stop real-time information displays for public transport: effects on customers. *Transportation Research Part A: Policy and Practice*, 41(6):489–501.

DAriano, A., Pacciarelli, D., and Pranzo, M. (2008). Assessment of flexible timetables in real-time traffic management of a railway bottleneck. *Transportation Research Part C: Emerging Technologies*, 16(2):232–245.

Einstein, A. (1905). Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen. *Annalen der physik*, 322(8):549–560.

Evans, H. and Skiles, G. (1970). Improving public transit through bus preemption of traffic signals. *Traffic Quarterly*, 24(4):531–543.

Federal Highway Administration (2016). Traffic volume trends: July 2016. Technical report, U.S. Department of Transportation.

Filar, J. A., Manyem, P., and White, K. (2001). How airlines and airports recover from schedule perturbations: a survey. *Annals of operations research*, 108(1-4):315–333.

Freyss, M., Giesen, R., and Muñoz, J. C. (2013). Continuous approximation for skip-stop operation in rail transit. *Transportation Research Part C: Emerging Technologies*, 36:419–433.

Gordon, P. and Richardson, H. W. (1996). Beyond polycentricity: the dispersed metropolis, los angeles, 1970-1990. *Journal of the American Planning Association*, 62(3):289–295.

Gordon, P., Richardson, H. W., and Wong, H. L. (1986). The distribution of population and employment in a polycentric city: the case of los angeles. *Environment and Planning A*, 18(2):161–173.

Hadas, Y. and Ceder, A. A. (2010). Optimal coordination of public-transit vehicles using operational tactics examined by simulation. *Transportation Research Part C: Emerging Technologies*, 18(6):879–895.

Hall, R., Dessouky, M., and Lu, Q. (2001). Optimal holding times at transfer stations. *Computers & industrial engineering*, 40(4):379–397.

Hall, R. W. (1985). Vehicle scheduling at a transportation terminal with random delay en route. *Transportation Science*, 19(3):308–320.

Hane, C. A., Barnhart, C., Johnson, E. L., Marsten, R. E., Nemhauser, G. L., and Sigismondi, G. (1995). The fleet assignment problem: Solving a large-scale integer program. *Mathematical Programming*, 70(1-3):211–232.

He, Q., Head, K. L., and Ding, J. (2011). Heuristic algorithm for priority traffic signal control. *Transportation Research Record: Journal of the Transportation Research Board*, 2259(1):1–7.

He, Q., Head, K. L., and Ding, J. (2014). Multi-modal traffic signal control with priority, signal actuation and coordination. *Transportation Research Part C: Emerging Technologies*, 46:65–82.

Hickman, M. D. and Wilson, N. H. (1995). Passenger travel time and path choice implications of real-time transit information. *Transportation Research Part C: Emerging Technologies*, 3(4):211–226.

Holroyd, E. and Scraggs, D. (1966). Waiting times for buses in central london. *Traffic Engineering and Control*, 8(3):158–160.

Jacobson, J. and Sheffi, Y. (1981). Analytical model of traffic delays under bus signal preemption: theory and application. *Transportation Research Part B: Methodological*, 15(2):127–138.

Janos, M. and Furth, P. (2002). Bus priority with highly interruptible traffic signal control: simulation of san juan's avenida ponce de leon. *Transportation Research Record: Journal of the Transportation Research Board*, 1811:157–165.

Jiang, H. and Barnhart, C. (2009). Dynamic airline scheduling. *Transportation Science*, 43(3):336–354.

Kanai, S., Shiina, K., Harada, S., and Tomii, N. (2011). An optimal delay management algorithm from passengers viewpoints considering the whole railway network. *Journal of Rail Transport Planning & Management*, 1(1):25–37.

Knoppers, P. and Muller, T. (1995). Optimized transfer opportunities in public transport. *Transportation Science*, 29(1):101–105.

Lan, S., Clarke, J.-P., and Barnhart, C. (2006). Planning for robust airline operations: Optimizing aircraft routings and flight departure times to minimize passenger disruptions. *Transportation science*, 40(1):15–28.

Lee, K. K. and Schonfeld, P. (1991). Optimal slack time for timed transfers at a transit terminal. *Journal of Advanced Transportation*, 25(3):281–308.

Lo, S.-C. and Chang, W.-J. (2012). Design of real-time fuzzy bus holding system for the mass rapid transit transfer system. *Expert Systems with Applications*, 39(2):1718–1724.

Lohatepanont, M. and Barnhart, C. (2004). Airline schedule planning: Integrated models and algorithms for schedule design and fleet assignment. *Transportation Science*, 38(1):19–32.

Ma, W., Liu, Y., and Yang, X. (2013). A dynamic programming approach for optimal signal priority control upon multiple high-frequency bus requests. *Journal of Intelligent Transportation Systems*, 17(4):282–293.

Ma, W., Yang, X., and Liu, Y. (2010). Development and evaluation of a coordinated and conditional bus priority approach. *Transportation Research Record: Journal of the Transportation Research Board*, 2145:49–58.

Milkovits, M. (2008). Modeling the factors affecting bus stop dwell time: use of automatic passenger counting, automatic fare counting, and automatic vehicle location data. *Transportation Research Record: Journal of the Transportation Research Board*, 2072:125–130.

Nesheli, M. M. and Ceder, A. A. (2014). Optimal combinations of selected tactics for public-transport transfer synchronization. *Transportation Research Part C: Emerging Technologies*, 48:491–504.

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

Newell, G. (1977). Unstable brownian motion of a bus trip. In *Statistical Mechanics and Statistical Methods in Theory and Application*, pages 645–667. Springer.

Niu, H. (2011). Determination of the skip-stop scheduling for a congested transit line by bilevel genetic algorithm. *International Journal of Computational Intelligence Systems*, 4(6):1158–1167.

Perk, V., Flynn, J., and Volinski, J. M. (2008). Transit ridership, reliability, and retention. Technical report, National Center for Transit Research and Research and Special Programs Administration and Florida Department of Transportation.

Petersen, J. D., Sölveling, G., Clarke, J.-P., Johnson, E. L., and Shebalov, S. (2012). An optimization approach to airline integrated recovery. *Transportation Science*, 46(4):482–500.

Schaafsma, A. and Bartholomeus, M. (2007). Dynamic traffic management in the schiphol bottleneck. In *Proceedings of the Second International Seminar on Railway Operations Research*, Hannover, Germany.

Schöbel, A. (2001). A model for the delay management problem based on mixed-integer-programming. *Electronic notes in theoretical computer science*, 50(1):1–10.

Seward, S. and Taube, R. (1977). Methodology for evaluating bus-actuated, signal-preemption systems. *Transportation Research Record: Journal of the Transportation Research Board*, 630(1):11–17.

Skabardonis, A. (2000). Control strategies for transit priority. *Transportation Research Record: Journal of the Transportation Research Board*, 1727:20–26.

Sweet, M. N., Bullivant, B., and Kanaroglou, P. S. (2016). Are major canadian city-regions monocentric, polycentric, or dispersed? *Urban Geography*, pages 1–27.

Thiele, T. N. (1880). Om anvendelse af mindste kvadraters methode i nogle tilfælde, hvor en komplikation af visse slags uensartede tilfældige fejlkilder giver fejlene en systematiskkarakter. *Det Kongelige Danske Videnskabernes Selskabs Skrifter-Naturvidenskabelig og Mathematisk Afdeling*, pages 381–408.

Tirachini, A. and Hensher, D. A. (2011). Bus congestion, optimal infrastructure investment and the choice of a fare collection system in dedicated bus corridors. *Transportation Research Part B: Methodological*, 45(5):828–844.

Visentini, M. S., Borenstein, D., Li, J.-Q., and Mirchandani, P. B. (2014). Review of real-time vehicle schedule recovery methods in transportation services. *Journal of Scheduling*, 17(6):541–567.

Von Smoluchowski, M. (1906). Zur kinetischen theorie der brownschen molekularbewegung und der suspensionen. *Annalen der physik*, 326(14):756–780.

Watkins, K. E., Ferris, B., Borning, A., Rutherford, G. S., and Layton, D. (2011). Where is my bus? impact of mobile real-time information on the perceived and actual wait time of transit riders. *Transportation Research Part A: Policy and Practice*, 45(8):839–848.

Wiener, N. (1923). Differential-space. *Studies in Applied Mathematics*, 2(1-4):131–174.

Zlatkovic, M., Stevanovic, A., and Martin, P. T. (2012). Development and evaluation of algorithm for resolution of conflicting transit signal priority requests. *Transportation Research Record: Journal of the Transportation Research Board*, 2311(1):167–175.