Title
Improving Bus Service with Dynamic Holding Control: from Theory to Practice

ISBN
9781369841084

Author
Argote-Cabanero, Juan

Publication Date
2014-12-01

Peer reviewed

Improving Bus Service with Dynamic Holding Control: from Theory to Practice

by

Juan Argote Cabanero

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

in

Engineering - Civil and Environmental Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Carlos F. Daganzo, Chair
Professor Samer Madanat
Professor Rhonda Righter

Fall 2014

Improving Bus Service with Dynamic Holding Control: from Theory to Practice

Abstract

Improving Bus Service with Dynamic Holding Control: from Theory to Practice

by

Juan Argote Cabanero

Doctor of Philosophy in Engineering - Civil and Environmental Engineering

University of California, Berkeley

Professor Carlos F. Daganzo, Chair

Service unreliability is widely recognized as one of the main deterrents for travelers to use buses as their primary mode of transportation. Bus systems are exposed to an adverse feedback loop that generates a tendency for buses to fall out of sync. This tendency can be counteracted by the application of control strategies that regulate the motion of the buses. Even though this is well known among transit operators and research has been devoted to solve this issue, in practice most bus systems rely on static control methods that ignore the real conditions of the system. Unfortunately, such an approach slows buses down excessively and it cannot compensate for severe disruptions.

This research proposes a control method that applies dynamic holding based on real-time conditions, in order to synchronize the buses' motion and improve their service reliability without imposing an excessive burden on their speed. A formulation that improves and simplifies existing dynamic holding control strategies for isolated bus lines is developed. Stability conditions both for headway and schedule deviations are then derived, and the theoretical results are validated with a simulation case study. The improved formulation is then extended to corridors where multiple bus lines overlap, reproducing the stability results found in the single line scenario. A framework for a real application entailing a distributed system of on-board devices is also presented. The control performance is then validated, now using a real-world case study. Finally, the robustness of this control framework is assessed considering the effects of possible device failures. In particular, the effects of GPS malfunctions on the on-board devices are studied analytically, revealing the resilience of the proposed control strategy.

A mis padres, por su apoyo incondicional.

Contents

List of Figures

List of Tables

Acknowledgments

Five years ago, when I embarked on my graduate studies, I did not anticipate how much this experience would change me. Learning from and collaborating with UC Berkeley's faculty and interacting with my fellow students has given me a brighter perspective on the academic environment.

Carlos Daganzo, my adviser, has been a constant source of inspiration with his passion, brilliance, and humility. I am deeply grateful for his continued support and guidance. Carlos has taught me to focus on the right questions and always look for the bigger picture when solving a problem.

I am also happy to thank all of the ITS faculty and student body for providing an engaging and collaborative environment where we all thrived. I have had the opportunity to directly work with many different professors and students: Alex Skabardonis, Alex Bayen, Samer Madanat, Ethan Xuan, Eleni Christofa, Vikash Gayah, Sebastién Blandin, Dan Work, and Aldo Tudela. In all cases, it has become clear why ITS is widely regarded as the world's leading institution in transportation engineering.

I also must thank the La Caixa scholarship and the Federal Highway Administration Dwight D. Eisenhower Fellowship program for their funding and support. Without their endorsement this work would have not been possible. Being a fellow has allowed me to be a part of an amazing group of individuals. Among them, Ricardo Cervera and Carlos Fernández with whom I have developed close bonds during all these years in the Bay Area.

I also want to mention the role that Carlos, Ethan, Dylan Saloner, and Jacob Lynn have played during the past years. Together we have created VIA Analytics, a startup project. VIA has become the vehicle to implement some of the ideas in this work. It has showed me how challenging and, at the same time, how rewarding it can be to see research implemented in the real world. I will forever treasure the time and experiences shared with my colleagues on this project.

In addition, I would like to thank my other friends. While pursuing my doctoral degree I have lived with Mike, Mal, Eric, and Alex, and in all cases I have had the greatest possible time. I am also grateful to Jean for becoming one of my best friends and teaching me a lesson or two in the process. I would also like to thank my friends on the other side of the Atlantic ocean: Edu, Nora, Mariona, Agus, Andres, Dani, and many more. Their sole presence made going back home the best time of the year. I also want to acknowledge my intramural basketball teammates for allowing me to be a part of two different IM dynasties at Berkeley's Recreational Sports Facility (RSF).

Fredrika, I thank you for, as you say, distracting me from my dissertation and making me a better person every day. Finally, I am infinitely indebted to my parents. Their perseverance, encouragement, and support made me who I am today.

As Carlo would learn, in the end I also realized that "the journey is the treasure".

Chapter 1

Introduction

The steady rise of congestion issues in urban environments [65, 34] and the spread of environmental awareness in modern societies [64] have made public transportation systems a more relevant mobility option worldwide. These systems, when correctly designed and operated, have proven to be efficient and practical alternatives to the private vehicle. Public transportation networks provide more equitable mobility and allow for the consolidation in time and space of the number of trips people make. The economies of scale that public transportation networks present [13] generate a series of advantages: lower emissions levels, greater space efficiency and, in some cases such as bus systems, relatively small capital and operational costs.

However, these advantages are conditioned by the ridership levels. If public transportation systems carry a low number of passengers, several negative consequences unfold [19]. In those cases, the operating costs and emission rates per passenger could rise above those associated with the private vehicle. To prevent this scenario and make public transportation an attractive alternative to other transportation modes it is important to know which are the service characteristics that users value most.

Several studies have shown that travel time and service reliability are generally regarded as the two most critical attributes [39, 75, 59, 73, 49]. Moreover, recent research has shown that public transit passengers may be discouraged from using public transportation if they consistently suffer unreliable service or even if they experience a single but very dramatic unreliability episode [18].

But service unreliability does not only have a pernicious effect on transit users, increasing their average travel times and generating stress. It also affects the costs of public transportation agencies. For example, if due to congestion or other effects, the headway or inter arrival time between consecutive buses becomes excessively large, transit agencies may have to resort to dispatching an additional bus to cover the delayed bus service. This means that agencies must backup drivers, resulting in an increase of operating costs. Thus, it is clear that providing reliable service should be an essential priority for any public transportation agency.

1.1 Problem Description

Unfortunately bus lines with high frequencies and numerous stops are naturally unstable. Buses face an adverse feedback loop that results in a tendency to fall out of schedule and get paired up. This in turn is reflected as an inherent lack of reliability.

As first explained in [57], the bus dwell times at the stops increase with the number of passengers waiting to board. Therefore if a bus suffers a disruption and gets delayed, it will have a greater chance of encountering more passengers waiting, falling even further behind.

Conversely, the follower bus will be likely to encounter fewer passengers, since its arrival interval with respect to the preceding bus will now be shorter than planned. Thus, the follower bus will naturally start running early. The drifting process will continue until the two buses pair. This phenomenon, popularly known as bus bunching (depicted in Figure 1.1), is perceived by bus users as a clear sign of service unreliability.

Figure 1.1: Bus bunching example, Berkeley, CA.

Different methods can be used to alleviate bus bunching. The work in this dissertation focuses on a particular family of control methods, known as holding control. They are based on the insertion of slack time in the planned bus operations [see an early example in 58]. The extra time is then used to compensate for the disturbances that buses encounter along the route, counteracting the natural tendency of the service to degrade.

1.2 Dissertation Goals

The objective of this dissertation is threefold. First, it attempts to provide a general modeling framework for holding control methods. Multiple holding methods have been proposed in the past but no one has previously attempted to generalize their formulation. The proposed framework is used as a ground for comparison of various new and existing holding control methods, providing stability analysis of different performance metrics such as schedule deviations or headway variability. This formulation is initially defined under the assumption of an isolated bus line, in alignment with the work on bus control that exists to date.

The second dissertation goal is to expand the scope of holding control methods to scenarios where multiple lines overlap in a single corridor. In real transit systems, corridors where different bus lines overlap are common. In those situations, control methods proposed for a single line, which ignore the interactions between the different lines, are bound to fail in improving the service and guaranteeing its stability. Expanding the holding control methods to corridors with overlapping lines will extend the control's applicability range to a multitude of transit agencies, helping to make bus systems a competitive alternative versus the private vehicle.

The final goal of this dissertation is improving our understanding of some real-world implementation issues. In the past, technological limitations have constrained the spread of holding strategies based on continuous real-time monitoring of the global state of the system. The sustained improvements and price reductions of mobile hardware and communications have lead to a scenario where it is now possible to apply holding control methods using off-the-shelf commoditized hardware. This dissertation aims to explore these new possibilities. A potential system architecture is developed and multiple real-world issues that could affect the control performance are also studied.

1.3 Dissertation Outline

This dissertation is structured as follows. Chapter 2 reviews the existing literature on bus service control methods. A special emphasis is placed on holding control methods. This chapter also examines the few case studies that describe implementations of holding control methods in the real world. Finally, the literature on the design and operation of multi-line bus systems is also summarized.

Chapter 3 presents the modeling framework used to generalize the formulation of holding methods in isolated lines. All the different assumptions used are described in detail and justified. Stability results are then derived for different classes of holding methods.

Chapter 4 extends the formulation and illustrates how the control can be calibrated in a real bus line with heterogeneous noise and demand. It includes a simulation-based case study using data collected from the UC Berkeley Campus Shuttle system comparing how different methods can improve service.

Chapter 5 extends the dynamic holding formulation for a single line to scenarios where multiple lines overlap. This chapter shows that by considering the right parameters in the control, it is possible to make multiple overlapping lines behave as if their buses were working in isolation, easily extending the stability conditions which were previously derived.

The multi-line control is then evaluated in a case study implementation in San Sebastian, Spain. The system is initially simulated and those results are then compared with empirical data collected from the field. This chapter concludes by describing the different issues that could affect the performance of the control in a real setting. One of the possible malfunctions, intermittent GPS availability, is studied in detail. Analytic expressions are derived to model the motion of an affected bus and several simulations are performed to determine the severity of this effect and the resilience of the control when recovering from a GPS loss.

The final chapter discusses potential extensions of this work and summarizes its contributions to the public transportation control literature.

Chapter 2

Literature Review

This chapter provides an overview of the different types of control methods that have been proposed to improve bus service. In particular, it focuses on the different holding control methods, because they are most directly related to this research. The chapter then presents different case-studies that have recently applied some of these strategies in the field. Finally, the design and operation of bus networks are also briefly reviewed. This is relevant since it reveals the type of scenarios in which expanding holding control methods to multiple lines can be beneficial.

2.1 Bus Control Methods

Bus control methods are generally classified under three different categories [see 31]: station control, inter-station control, and other ad-hoc measures (e.g. vehicle addition). This classification differentiates control methods by considering the location where they are applied. Station control methods generally include holding methods, where buses are retained for short periods of times at one or multiple locations known as control points, and boarding limits methods, which constrain the number of users that can board a particular bus at a stop. On the other hand, inter-station control methods are generally based on exerting speed control.

It is possible to show that speed control methods can be reformulated to be equivalent to station-based holding methods.[1] For this reason, it is proposed in this review to categorize the control methods by the form of control: either holding or skip-stop/boarding-limits. Instead of considering how the control is spatially applied as the differentiating factor, the strategies are now classified by their nature. While the first family affects the cruising motion of the bus, the second family affects the boarding/alighting process.[2]

[1] Operating speed control methods can be referred to as Lagrangian methods while station-based holding methods are their Eulerian counterpart. The connection between both types of specifications in the context of vehicle motion is explained in [24].

[2] Transit agencies sometimes resort to other operational practices to deal with instabilities in a bus line

2.1.1 Boarding-Limits and Skip-Stops

Stop skipping [as studied in 52, 68, 35, 70] can be used as a control method to improve the reliability of the bus operation. This method can effectively increase the operating speed of a delayed bus by not serving some of the stops along the route. At first sight, this may seem like an adequate method. However, with such method, the passengers waiting at a skipped stop have to wait an additional headway until the next bus arrives. Moreover, passengers wanting to alight at the skipped stop will have to transfer to another bus or walk an extra distance. For these two reasons, stop skipping is not systematically applied by transit agencies as a mean to improve reliability. In this context, it is solely used as a last resort, to decouple two buses that have bunched.

However, stop skipping is sometimes used by transit agencies for other purposes such as deadheading, expressing and short-turning [see 37, 51]. In those cases, the stops that are skipped are planned in advance. The goal of these practices is to boost the frequency or speed of the bus service in particular segments of a line. Clearly informing passengers ahead of time of the distinct types of service is essential for the success of these practices [20].

Another possible approach to improve bus service reliability consists in limiting the number of boardings under certain conditions. This approach is known as boarding limits. It is not as invasive as stop skipping because on-board passengers are allowed to alight.

However, even though this option may work well in theory, it is difficult to implement in practice. Boarding limits are difficult to enforce without a dedicated workforce or an expensive infrastructure that could guarantee the queue discipline during the boarding process. Under this sort of control, passengers have perverse incentives to try to cut in line and advance their boarding order. Perhaps this is the reason why this type of method, contrary to stop skipping, has not been extensively covered in the literature. A recent article [27] studied this control method combined with holding methods. However, the group that developed this work later implemented a pilot project where only holding control was applied [54].

2.1.2 Holding Methods

The various holding methods proposed in the literature follow a common principle. They all introduce slack time in the planned motion of the buses. This additional time is then used to modulate the pace of the vehicles and counteract the effects that the stochasticity of traffic and demand have on service regularity.

The adjustment of the vehicles' pace can be achieved in two ways. One option is holding buses when necessary at specific locations (generally termed control points) for a dynamically determined time. This is the most widely used strategy since it is fairly simple to apply. In practice, a bus operator can precisely match the period in which the bus is stopped with the control method's recommendation.

---

such as adding vehicles. However, this practice does not directly affect the mechanism behind the bunching phenomenon and simply delays its appearance.

The other option is to provide speed recommendations to the drivers so that they can adapt their pace while cruising between stops. This option is the most appropriate for transportation modes with dedicated infrastructure like rail. However, it is much more challenging to implement in real bus systems. Buses do not necessarily operate in segregated lanes and, even when they do, their cruising motion is affected by external elements such as traffic signals or turning vehicles. Thus, bus operators will in most cases have difficulty precisely following speed recommendations.

Leaving aside the physical implementation of the control, the two key questions that characterize holding methods are: how much slack needs to be introduced in the schedule? And how should the dynamic holding times or recommended speeds be calculated? The holding methods covered in the next paragraphs differ from one another in how these two questions are approached.

Traditional schedule-based control Traditional schedule-based control is normally applied in low frequency bus services. In those cases, keeping buses on schedule is extremely important because users arrive at the stop following on the published bus schedule time [22]. In this method the holding recommendation for a given bus is determined using a target scheduled time. If a bus arrives to the control location before the scheduled time, it is held until the scheduled time is reached. Otherwise, the bus is dispatched immediately after all boardings and alightings are completed.

The slack in the scheduled times is determined based on an offline analysis of the buses' historical performance. The transit agency generally sets the slack so that it is possible to compensate for all possible disturbances. Since this is a static approach that ignores real-time conditions, large amounts of of slack are usually required. For this reason, schedule-based control improves reliability, but at the price of significantly impacting the service speed. Despite this drawback, it is the method most widely used in practice because of its simplicity.

Headway-based control When bus service is very frequent, some operators resort to what is known as headway-based operation. The notion of a schedule is substituted by a target inter-arrival time between consecutive buses. This is a typical approach for high frequency bus services. Under short headways (<10 minutes), passengers do not rely on a schedule and arrive randomly at the stations. Reliability is then directly related to the variability of the bus headways, and the control methods aim to keep regular inter-arrival times.

Headway-based methods use headway values as inputs, rendering any sort of underlying schedule unnecessary for its application. This type of control was first analyzed in [60], following a control form based on minimum headway thresholds instead of schedule slack. This pioneering work analyzed a very simplistic and idealized system with a single control location and only two vehicles. The paper describes a dynamic programming formulation that minimized total passenger wait time, yielding a parametric expression of the optimal

headway threshold. At the control point, only those buses whose headway was less than the threshold value are held.

Later references in the early literature examine optimal threshold headway control policies in more complex scenarios. For example, [10] analyzes a two-terminal service with a control location in between and multiple vehicles. In [58], the analysis framework is expanded by separating the vehicles' cruising and loading times in the formulation. Additional references such as [72, 36] have examined other variants of the threshold-based control.

Unfortunately, the thershold method is difficult to analyze for systems with many buses and control points. Thus, even though the control strategy is rather simple, it has not been optimized for systems of realistic size.

Headway-based control without rigid thresholds More complex methods use the real-time state of a bus line to determine holding recommendations. In view of the complexity of the problem, the initial work in this area used simulation to analyze the problem. As mentioned in [45], the stochastic nature of bus operations and the difficulties to describe them analytically make that approach attractive. Some important references that use simulation to examine headway-based holding are: [47], [2], [1], [74], [66] and [3]. The common conclusion in these papers is that headway-based methods are more effective than schedule control and headway-based threshold holding, although both can significantly improve the reliability in a bus corridor with a single bus line.

The emergence of Global Positioning System (GPS) technology, which allows tracking vehicles precisely in real time, enabled the application of even more complex forms of control. An initial example that incorporates real-time GPS information into the holding problem can be found in [30]. However, this reference considers only the minimization of out of vehicle waiting times, ignoring in-vehicle passenger delay. A similar problem is studied in [69], but this time in-vehicle delay is included in the objective function. In both cases, it is shown that applying headway-based holding control at multiple stations can achieve better results than applying the control at a single location. However, both examples are based on the formulation of a complex mathematical program that requires the use of heuristics in order to obtain the control recommendations.

On the other hand, [23] proposed a modeling approach that linearized the state equations of an isolated bus line with multiple control points. This approximation allows a parametric analysis of the holding problem. This work shows that under certain conditions it is possible to obtain closed form solutions for the optimal control policies. The resulting control minimizes the average delay buses suffer en route in order to prevent bunching and counteract the effect of small disturbances.

An expansion of this method which is able to deal with large disruptions can be found in [26]. This method, instead of applying holding recommendations at control points, regulates the speed of a given bus based on the expected demand in the line and the current spacings between the bus and its two contiguous buses. The resulting headway-based dynamic holding method in these two references can be implemented in a decentralized manner. This is

useful because it facilitates the deployment of the control in the real-world in a scalable way, leveraging the latest developments in mobile technologies.

Finally, [11] proposes an alternative method that ignores any predefined headway. It implements a simple holding rule based on real-time arrivals to a control location. It allows the lines' natural headway to emerge while regularizing the spacing of the buses. This method has also been shown to avoid bus bunching. However, it is not well suited for the operation of bus lines that need to guarantee service frequencies.

This dissertation uses the promising modeling approach in [23] and expands it to consider the control of schedule-based operations and guarantee its stability. Moreover, because all the references reviewed in this section are limited to the control of isolated bus lines a later chapter of this dissertation will consider the simultaneous control of multiple lines.

2.2 Existing Case Studies

As we saw in the previous section, there exists a wealth of literature that theoretically approached the bus control problem. However, almost none of the proposed control methods have been tested in a real setting due to the difficulties associated with such endeavor. In some instances, researchers have resorted to real data gathered with Automated Vehicle Location (AVL) [28, 36] or other sources such as fare-card information [50] to assess reliability levels on certain routes and calibrate complex simulation frameworks.

These examples, even though they incorporate more realism into their analyses, are still limited. Simulation frameworks, just like their theoretical counterparts, are based on a set of assumptions. These assumptions may capture the essence of the real operation but they cannot perfectly reflect the complexity associated with a real implementation. For example, a simulation may assume that bus drivers will follow the control recommendations perfectly, while in reality, for a multitude of possible reasons, drivers may not do so.

Examples of real-time dynamic bus control implementations are scarce. The first reported test of this sort of control in a real bus system appears to be [11]. This reference applied a method to equalize headways in a university campus route. Students located in two different control points noted the buses' arrival times and combined this information with arrival predictions for the following bus in order to obtain holding time recommendations on the spot. Although this approach would be difficult to apply on a daily basis in a real system, it produced considerable improvements in headway regularity.

Other practical implementation of dynamic real-time holding control can be found in [54]. This work shows the preliminary results of an application of a version of the control methodology developed in [27] that only considers holding as a control method. This control was implemented in one of the main lines of Transantiago, the transit operator in Santiago, Chile. The preliminary testing presented in this reference was also carried out using individuals who were physically assigned to each one of 24 control locations. Since [27] requires complex numerical procedures to solve a rolling horizon optimization problem, the recommendations were in this case computed on a central server. The holding recommendations

were then communicated to these persons via sms and relayed to the drivers at the control locations. The reference also describes a more advanced technological implementation that is underway in which the holding information will be communicated to the drivers using an industrial console that is already integrated in the buses' dashboard.

Neither of the examples just described were ready for deployment in a large real-life setting, because the level of automation was rudimentary. This dissertation will provide results from the first completely automated implementation of dynamic control, and will do so for multiple coordinated bus lines.

2.3 Overlapping lines literature

To the author's knowledge, no bus control reference has previously considered the operation of buses in corridors with overlapping lines. However, these scenarios are common in reality. This becomes apparent by glancing over any map of bus service in a major city in the US. For example, all of the three major bus networks that serve the San Francisco Bay Area (Muni, AC Transit, and VTA) present multiple corridors in their networks where several routes overlap.

(a) AC Transit's network: Lines 7 and 18 overlap at Shattuck Ave.

(b) NY MTA's network: Lines 1, 3 and 4 overlap at 5th Ave., among others.

Figure 2.1: Real corridors with overlapping lines.

Even though the bus control literature had so far neglected these scenarios, it is possible to find some examples within other areas in the public transportation literature that address the multi-line corridor scenario. These areas can be divided into those concerning bus service design and those related to the operation of such service. We will see that in both cases there is a tendency to take a certain relability level for granted, even though this is something that existing control methods are not well equipped to do for multi-line corridors.

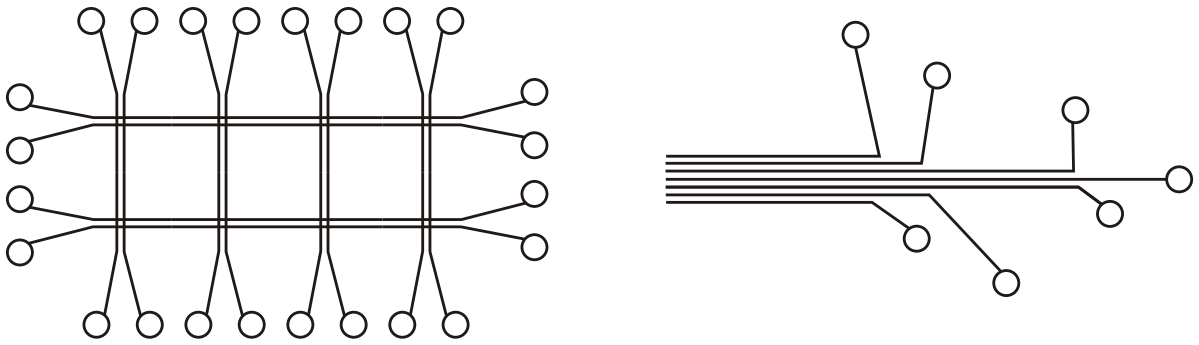
Design

The design of bus service generally does not directly involve service reliability. Instead it is assumed that a certain level of reliability is achieved. Thus, service reliability is normally neglected from the design formulation, even though the proposed corridor or network configuration may be difficult to control (e.g. due to the provision of overlapping bus lines).

The network design problem is normally approached in one of two ways. The most studied approach consists in solving a combinatorial optimization problem to identify the routes that minimize a generalized cost function, generally including agency and user cost [e.g. 9, 61]. Although it is well known that service reliability has a great influence on user cost, reliability is rarely included in the formulation of the objective function.

The second approach to network design is based on the formulation of continuous models. The continuous approximations also consider both the agency and user costs. For the sake of simplicity, these continuous models also assume a given level of service reliability. However some of these models assume a network configuration where overlapping bus lines play a fundamental role. For instance [25], using continuous approximations, analyzed the design of a hybrid network structure. This network configuration presents a grid structure in the center combined with a hub and spoke structure in the periphery, as illustrated in Figure 2.2a.

Other examples of network configurations that include the presence of overlapping lines are trunk and branch systems (Figure 2.2b). In this case, several bus lines which cover different regions in the periphery of the urban area converge in a single corridor to access the CBD. Real examples of this type of configuration can be found currently in Ottawa, Canada, and Brisbane, Australia.



(a) Hybrid network. (b) Trunk and branch.

Figure 2.2: Network configurations that include overlapping lines.

All these corridor and network configurations would clearly benefit from the implementation of control methods that could guarantee a reliable service.

Operation

There are various examples within the bus operation literature that relate to overlapping lines. One clear case is the study of bus stop capacity. This literature tries to assess the rates at which buses can discharge from a bus stop under different operating conditions. Reference [42] provides an insightful analysis on the capacity of curbside bus stops. One of its main findings is that the variation of bus arrival times to a stop can negatively affect its capacity. Therefore, providing regular service in corridors with frequent bus arrivals (as will be the case when multiple lines operate simultaneously) is important to improve the performance of the stops.

Another context that involves overlapping routes is the bus allocation problem [e.g. 44, 41].Considered is an existing route configuration where several lines coincide. The objective is to optimally allocate a given number of buses among these routes so as to minimize user cost. The problem assumes that the service frequency remains constant, ignoring again the challenge of providing reliable service in those systems. Thus, the use of control methods is required to obtain the results predicted by these optimization procedures.

Chapter 3

Dynamic Holding Control for an Isolated Line*

This chapter first presents the model used to characterize the motion of the buses in an isolated line. The proposed control method is based on the concept of a so-called virtual schedule. The control formulation applies both to schedule-based and headway-based bus services.[1] Both control methods are expressed as particular cases of a general linear control formulation. The stability of this generalized linear control is then analyzed. These results are then used to obtain a control rule that maximizes bus commercial speed under reliability constrains. Finally, the performance of different control rules is assessed with simulation using realistic data.

3.1 Isolated Line Model

3.1.1 Bus Service Assumptions

We consider a line of unbounded length in which buses are indefinitely dispatched from the origin station with constant and regular headways. This scenario is comparable to real-world bus lines with a finite number of buses that have enough layover time to guarantee that the origin station always has a pool of buses ready to be dispatched. We will further assume that the bus capacity is unlimited, which is reasonable if a transit agency has considered the fleet and vehicle sizing problems in order to decide their target dispatching intervals.

Cruising The motion of an individual bus between stations will be characterized using the following assumptions: (i) buses are assumed to not pass each other and (ii) the average

*Much of this chapter is based on [77]. The contributions of my co-authors are duly noted and acknowledged.

[1] A publishable schedule can be obtained by shifting the virtual schedule adequately, while in the case of headway-based operations the virtual schedule would not be accessible to the public.

travel times between stations are assumed to be known, but they are time and location dependent. Assumption (i) is reasonable if the controlled system is stable so that buses don't exhibit a tendency to fall out of sync. Moreover, this particular assumption can also hold in urban corridors with high traffic, where buses may be constrained to remain in their lane.[2] Assumption (ii) reflects the operation of modern bus services equipped with Automated Vehicle Location (AVL) capabilities [46], where time and space-dependent average travel times are easily accessible.

Dwelling The bus dwelling process, on the other hand, is characterized by these additional assumptions: (iii) passenger arrival rates are known and the depend on the station and time of day, (iv) the mean bus loading time is predominantly affected by boarding passengers and is therefore proportional to the elapsed time between consecutive bus arrivals, (v) the holding times are applied immediately after the boarding process is concluded and only those passengers arriving in the inter-arrival period of time would be allowed to board the bus, (vi) buses do not skip stops and holding is applied at all stations, and (vii) there exists enough slack in the virtual schedule to ensure that the holding time never runs short.

Assumption (iii) applies in modern transit systems, where passenger demand can be precisely monitored using Automated Passenger Counting systems or fare card data [38, 62]. Assumption (iv) is appropriate for bus systems where passengers board and alight simultaneously and the alighting time per passenger is much smaller than the boarding time per passenger. Assumption (v) greatly simplifies the formulation with a negligible effect, since loading and holding times are much shorter than the inter-arrival time. Assumption (vi) can be relaxed by simply considering the uncontrolled motion between adjacent control stations, something that it is also explored in this work. Lastly, assumption (vii) allows the formulation to be linearized and makes the problem tractable. This will be shown in the next section. The section will also show how to relax this assumption by adjusting some coefficients in the model.

Noise Finally, (viii) the noise that affects the actual vehicle trip times between stations, including both cruising and loading times, is assumed to be generated by independent random variables with a variance that is independent of the headway. This assumption is a reasonable approximation if buses are operated without severe congestion, or with their own right-of-way, i.e., segregated from car traffic.

3.1.2 Formulation

Using the previous assumptions it is possible to formulate a succinct linear model with white noise that characterizes the motion of the buses in our idealized line. Let us use $s = 0, 1, \cdots$ as the station index, where $s = 0$ denotes the origin station and the index increases

[2]Passing could be allowed in reality, as long as the maneuver is correctly identified and reflected in the application of the control.

monotonically with step 1 per station in the direction of travel. The index $n = 0, 1, \cdots$ is used as the bus index, with $n = 0$ denoting the first bus dispatched. Given below, the rest of the notation uses this indexing approach:

- $t_{n,s}$ is the scheduled arrival time of bus n at station s. The $t_{n,s}$ form the virtual schedule for buses, which in general should not be the published schedule to passengers.[3]
- $a_{n,s}$ is the actual arrival time of bus n at station s.
- $\varepsilon_{n,s} = a_{n,s} - t_{n,s}$ is the deviation from schedule of bus n upon arrival to station s. A positive value of this parameter signifies that the bus arrived late while negative is associated with earliness.
- H is the target headway in the line. Since the headway is assumed constant it can also be expressed as $H = t_{n,s} - t_{n-1,s}$, $\forall n > 0$.
- $h_{n,s} = a_{n,s} - a_{n-1,s}$ is the actual headway between bus n and its leading bus, $n-1$, at station s.
- $c_s(t_{n,s}) \equiv c_{n,s}$ is the average cruising time from station s to $s+1$ for bus n, which includes the time to accelerate and decelerate, but does not include the dwell time to serve passengers.[4]
- $\nu_{s+1}(t_{n,s}) \equiv \nu_{n,s+1}$ is the random noise in the trip time of bus n between stations s and $s+1$, whose mean is zero and variance is $\sigma^2_{n,s+1}$.[4]
- $D_{n,s}$ is the actual holding time applied to bus n at station s. The value of this variable depends on the control method used and the state of the system.[5]
- $d_{n,s}$ is the amount of pre-planned slack time included in the virtual schedule at station s. This would be the applied holding time if the bus arrives on time.
- t_b is the average boarding time per passenger.
- $\beta_s(t_{n,s}) \equiv \beta_{n,s}$ is a unitless demand parameter. It can be described as the marginal loading time that buses experience, i.e. the change in total time that arises when the

[3] If desired, a published schedule can be obtained by shifting the virtual schedule earlier in time to ensure that buses never depart ahead of the published schedule.

[4] Note that this term is assumed to be time dependent. However, we also implicitly assume that conditions on the line do not change abruptly and that controlled buses will not deviate from their schedule excessively, so that the dependency can be made with respect to the planned arrival time.

[5] Note that in reality buses can be held at a station by delaying their departure after the passenger loading is completed or by dynamically adapting their speed en route. Both approaches present certain advantages. The first one is easy to implement precisely, since a bus can always remain stopped at a station. On the other hand, adapting the bus speed would be less noticeable to passengers and it also releases station capacity as well.

headway is increased by one unit. It can be obtained by normalizing the demand rate (in passengers/time) by the inverse of the average boarding time per passenger.[4,6]

Considering this notation the scheduled motion of the buses in the line can be obtained as: [7]

$$t_{n,s+1} = t_{n,s} + \beta_{n,s}H + d_{n,s} + c_{n,s}, \tag{3.1a}$$
$$t_{n,s} = t_{n-1,s} + H, \tag{3.1b}$$

which entails that the scheduled departure time for bus n at station s would be $t_{n,s}+\beta_{n,s}H+d_{n,s}$.[8] On the other hand, the actual arrival times can be expressed as:

$$a_{n,s+1} = a_{n,s} + \beta_{n,s}h_{n,s} + D_{n,s} + c_{n,s} + \nu_{n,s+1}. \tag{3.2}$$

By subtracting equation (3.1) from (3.2) one can obtain the expression for the deviations from schedule, $\varepsilon_{n,s}$, which are the state variables. A positive value of ε indicates arrival delay while a negative value indicates arrival earliness.

$$\begin{aligned}\varepsilon_{n,s+1} &= \varepsilon_{n,s} + \beta_{n,s}(h_{n,s} - H) + (D_{n,s} - d_{n,s}) + \nu_{n,s+1} \\ &= \varepsilon_{n,s} + \beta_{n,s}(\varepsilon_{n,s} - \varepsilon_{n-1,s}) + (D_{n,s} - d_{n,s}) + \nu_{n,s+1}.\end{aligned} \tag{3.3}$$

Let us now formulate the control expression, $D_{n,s}$, as a general linear function that depends on the schedule deviations of all buses in the line at station s.[9] This approach conveniently converts the state equation of the bus line into a linear equation with white noise, which facilitates the upcoming stability analysis. The chosen expression is:

$$D_{n,s} = d_{n,s} + \sum_i \gamma_i \varepsilon_{n-i,s}, \tag{3.4a}$$

which can be conveniently rewritten as:

$$D_{n,s} = d_{n,s} - [(1+\beta_{n,s})\varepsilon_{n,s} - \beta_{n,s}\varepsilon_{n-1,s}] + \sum_i f_i \varepsilon_{n-i,s}. \tag{3.4b}$$

[6]The following equality can help in understanding the parameter: $\beta_s H = \lambda_s H t_b$, where λ_s would be the demand rate at the stop. Typical values of $\beta_s(t_{n,s})$ range from 10^-2 in low demand scenarios to $2 \cdot 10^-1$.

[7]Note that for the sake of concision, the time dependency of the demand and average cruising time parameters will be denoted by including the bus subscript.

[8]Passenger boarding time is expressed $\beta_{n,s}H$ based on assumption (v).

[9]The schedule deviations of the current bus and its predecessors, $\varepsilon_{i,s}$, $i \in \{0, 1, \cdots, n\}$ should be available at the moment in which bus n arrives at station s. The schedule deviations of the buses behind, $\varepsilon_{i,s}$, $i \in \{n+1, \cdots\}$, can only be predicted. For now, we will assume that we have perfect information, i.e., we know all $\varepsilon_{i,s}$. It will be shown later that schedule deviation predictions are not really needed to obtain a near-optimal control method.

To verify, plug (3.4b) into (3.3) and note how the simple aforementioned linear equation with white noise is recovered:

$$\varepsilon_{n,s+1} = \sum_i f_i \varepsilon_{n-i,s} + \nu_{n,s+1}. \tag{3.4c}$$

Now note $f_i = \gamma_i$ for all i except $f_0 = \gamma_0 + 1 + \beta_{n,s}$ and $f_1 = \gamma_1 - \beta_{n,s}$. Also note that

$$\sum_i f_i = 1 + \sum_i \gamma_i. \tag{3.4d}$$

A specific holding control method under this approach will be fully determined by the choice of slack times $d_{n,s}$ at each station and the values of the different control coefficients $\{\cdots, f_{-1}, f_0, f_1, \cdots\}$. This formulation allows us to capture as special cases most of the proposed control methods in the literature.

Some Special Cases

First, note that under this framework uncontrolled motion corresponds to a scenario with control coefficients: $f_0 = 1 + \beta_{n,s}$, $f_1 = -\beta_{n,s}$, $f_i = 0 \, \forall i \notin \{0, 1\}$, and $d_{n,s} = 0 \, \forall s$. In this case, the resulting holding times and motion equations are:

$$D_{n,s} = 0, \qquad \text{(no control)} \tag{3.5a}$$

and

$$\varepsilon_{n,s+1} = (1 + \beta_{n,s})\varepsilon_{n,s} - \beta_{n,s}\varepsilon_{n-1,s} + \nu_{n,s+1}. \qquad \text{(no control)} \tag{3.5b}$$

The conventional schedule-based control can also be formulated as a special case of the generalized formulation. This control method instructs drivers not to depart from a station until the scheduled departure time. If the underlying schedule contains enough slack, as per assumption (vii), then the drivers can always depart the control station on schedule, i.e., $a_{n,s} + \beta_{n,s}h_{n,s} + D_{n,s} = t_{n,s} + \beta_{n,s}H + d_{n,s}$. Therefore, it follows that:

$$D_{n,s} = d_{n,s} - [\varepsilon_{n,s} + \beta_{n,s}(\varepsilon_{n,s} - \varepsilon_{n-1,s})], \qquad \text{(sch. control)} \tag{3.6a}$$

and

$$\varepsilon_{n,s+1} = \nu_{n,s+1}. \qquad \text{(sch. control)} \tag{3.6b}$$

The forward headway based method proposed in [23] is:

$$D_{n,s} = d_{n,s} - (\alpha + \beta_{n,s})(h_{n,s} - H), \qquad \text{(forw. headway)} \qquad (3.7a)$$

which can be expressed as a function of the deviations from schedule:

$$D_{n,s} = d_{n,s} - (\alpha + \beta_{n,s})(\varepsilon_{n,s} - \varepsilon_{n-1,s}). \qquad \text{(forw. headway)} \qquad (3.7b)$$

This is the special case of (3.4b) where $f_0 = 1 - \alpha$, $f_1 = \alpha$, and $f_i = 0 \, \forall i \notin \{0, 1\}$, and $0 \leq \alpha \leq 1$, so that the motion equation for the buses under this control method is:

$$\varepsilon_{n,s+1} = (1 - \alpha)\varepsilon_{n,s} + \alpha\varepsilon_{n-1,s} + \nu_{n,s+1}. \qquad \text{(forw. headway)} \qquad (3.7c)$$

Similarly, the Eulerian version of the (Lagrangian) method in [26], which depends on both the forward and backward headways, is the special case with $f_{-1} = f_1 = \alpha$, $f_0 = 1 - 2\alpha$, $f_i = 0 \, \forall i \notin \{-1, 0, 1\}$, and $0 \leq \alpha \leq 1/2$. The holding and motion equations in this case are:

$$D_{n,s} = d_{n,s} + \alpha(h_{n+1,s} - H) - (\alpha + \beta_{n,s})(h_{n,s} - H), \qquad \text{(two-way headway)} \qquad (3.8a)$$
$$D_{n,s} = d_{n,s} + \alpha\varepsilon_{n+1,s} - (2\alpha + \beta_{n,s})\varepsilon_{n,s} + (\alpha + \beta_{n,s})\varepsilon_{n-1,s}, \qquad \text{(two-way headway)} \qquad (3.8b)$$

and

$$\varepsilon_{n,s+1} = \alpha\varepsilon_{n+1,s} + (1 - 2\alpha)\varepsilon_{n,s} + \alpha\varepsilon_{n-1,s} + \nu_{n,s+1}. \qquad \text{(two-way headway)} \qquad (3.8c)$$

Finally, the backward headway control method used in [11] can be adapted to this framework, even though the original version does not require the existence of any underlying virtual schedule and it is completely independent of the demand levels. The basic backward-headway based control law originally proposed is $D_{n,s} = \alpha h_{n+1,s}$. If a virtual schedule with headway H is used, then it is possible to express the holding times as $D_{n,s} = \alpha H + \alpha(h_{n+1,s} - H) = \alpha H + \alpha\varepsilon_{n+1,s} - \alpha\varepsilon_{n,s}$.

Thus, in our framework, the slack time used by this particular control method would be αH, which can be quite large for the values of α recommended in [11] to maintain stability ($\alpha \approx 1/2$). To have a fair comparison with the other control options previously presented, the variant of the backward-headway based control that will be considered will have slack times $d_{n,s}$ and control parameter α which can be freely chosen. The resulting control method is obtained by setting $f_{-1} = \alpha$, $f_0 = 1 + \beta_{n,s} - \alpha$, $f_1 = -\beta_{n,s}$, and $f_i = 0 \, \forall i \notin \{-1, 0, 1\}$, yielding:

$$D_{n,s} = d_{n,s} + \alpha(h_{n+1,s} - H), \qquad \text{(backw. headway)} \qquad (3.9a)$$
$$D_{n,s} = d_{n,s} + \alpha\varepsilon_{n+1,s} - \alpha\varepsilon_{n,s}, \qquad \text{(backw. headway)} \qquad (3.9b)$$

and

$$\varepsilon_{n,s+1} = \alpha\varepsilon_{n+1,s} + (1 + \beta_{n,s} - \alpha)\varepsilon_{n,s} - \beta_{n,s}\varepsilon_{n-1,s} + \nu_{n,s+1}. \qquad \text{(backw. headway)} \qquad (3.9c)$$

The performance and stability of these different control methods will be assessed and compared both analytically and using a simulation framework.

3.2 Stability Analysis

The goal of this section is to determine which values of the control coefficients, f_i, guarantee the stability of the headways and deviations from schedule. The approach introduced in [23] is used.

Note that the state equation (3.4c) can be expressed in a simple form using the convolution operation (denoted with the symbol "$*$") of two vectors. The two vectors are the deviations from schedule of the buses arriving at station s, $\boldsymbol{\varepsilon}_s = [\cdots, \varepsilon_{n-1,s}, \varepsilon_{n,s}, \varepsilon_{n+1,s}, \cdots]^T$ and the set of control coefficients $\boldsymbol{f} = [\cdots, f_{-1}, f_0, f_1, \cdots]^T$. The n^{th} element of the convolution is the deterministic part of (3.4c): $[\boldsymbol{f} * \boldsymbol{\varepsilon}_s]_{n,s} = \sum_k f_{n-k}\varepsilon_{k,s} = \sum_i f_i\varepsilon_{n-i,s}$. If we also define the system noise in vector form as $\boldsymbol{\nu}_s = [\cdots, \nu_{n-1,s}, \nu_{n,s}, \nu_{n+1,s}, \cdots]^T$, then the vector form of (3.4c) is: $\boldsymbol{\varepsilon}_{s+1} = \boldsymbol{f} * \boldsymbol{\varepsilon}_s + \boldsymbol{\nu}_{s+1}$.

This convolution can be applied recursively going back to $s = 0$ so that the $\boldsymbol{\varepsilon}$ terms can be expressed as a function of only the control coefficients $\boldsymbol{f}$ and the noise terms $\boldsymbol{\nu}$. To do this write: $\boldsymbol{\varepsilon}_{s+1} = \boldsymbol{\nu}_{s+1} + \boldsymbol{f} * \boldsymbol{\varepsilon}_s = \boldsymbol{\nu}_{s+1} + \boldsymbol{f} * (\boldsymbol{\nu}_s + \boldsymbol{f} * \boldsymbol{\varepsilon}_{s-1}) = \boldsymbol{\nu}_{s+1} + \boldsymbol{f} * \boldsymbol{\nu}_s + \boldsymbol{f} * \boldsymbol{f} * \boldsymbol{\varepsilon}_{s-1} = \cdots$. Now define $\boldsymbol{f}_{|j}$ to be the j^{th} self-convolution of $\boldsymbol{f}$ (i.e., $\boldsymbol{f}_{|j} = \boldsymbol{f} * \boldsymbol{f}_{|j-1}$, where $\boldsymbol{f}_{|0} = [\cdots, 0, 1, 0, \cdots]^T$). Since it is assumed that buses are always dispatched from the first station on time ($\boldsymbol{\varepsilon}_0 = \boldsymbol{0}$), the above expression becomes:

$$\boldsymbol{\varepsilon}_{s+1} = \sum_{j=0}^{s} \boldsymbol{f}_{|\boldsymbol{j}} * \boldsymbol{\nu}_{s+1-j}, \qquad (3.10a)$$

which expands to

$$\varepsilon_{n,s+1} = \sum_{j=0}^{s} [\boldsymbol{f}_{|\boldsymbol{j}} * \boldsymbol{\nu}_{s+1-j}]_n = \sum_{j=0}^{s} \sum_i f_{i|j}\nu_{n-i,s+1-j}. \qquad (3.10b)$$

This expression will play a fundamental role in the stability analysis of both headway and schedule deviations. The latter is now taken up.

Schedule Adherence

The schedule deviation stability achieved by different control methods will be assessed by the variance of said deviation; i.e. by the function $\sigma^2_\varepsilon(\boldsymbol{f}, n, s)$ that returns $\text{var}(\varepsilon_{n,s})$ given

$\boldsymbol{f}$, n and s. Consider equation (3.10), since the noise terms are assumed to be independent and identically distributed (i.i.d.) with variance σ^2, we have:

$$\sigma_\varepsilon^2(\boldsymbol{f}, n, s) = \sigma^2 \sum_{j=0}^{s-1} \sum_i (f_{i|j})^2. \tag{3.11a}$$

Since all the terms in this summation are non-negative, an upper bound to the variance of $\varepsilon_{n,s}$ for any n and s is:

$$\sigma_\varepsilon^2(\boldsymbol{f}) \equiv \lim_{\substack{n\to\infty \\ s\to\infty}} \mathrm{var}(\varepsilon_{n,s}) = \sigma^2 \sum_{j=0}^{\infty} \sum_i (f_{i|j})^2 \geq \sigma_\varepsilon^2(\boldsymbol{f}, n, s), \tag{3.11b}$$

which is a measure of the control algorithm's effectiveness. It is now possible to establish the following.

Lemma 3.2.1. *Define* $F = \sum_i |f_i|$*, then* $\sum_i |f_{i|j}| \leq F^j$, $\forall j \geq 1$.

Proof. First, expand the convolution term as $\sum_i |f_{i|j}| = \sum_i |\sum_k f_{k|j-1} f_{i-k}|$. The absolute value of a sum of terms is smaller than the sum of the absolute value of those terms, thus: $\sum_i |\sum_k f_{k|j-1} f_{i-k}| \leq \sum_i \sum_k |f_{k|j-1}||f_{i-k}| = \sum_k (|f_{k|j-1}| \sum_i |f_{i-k}|) = F \sum_k |f_{k|j-1}|$. These steps can now be applied recursively for indices $\{j-2, j-3, \cdots, 0\}$ until reaching $\sum_i |f_{i|j}| \leq F^j \sum_k |f_{k|0}| = F^j$ □

Theorem 3.2.2 (Sufficient condition for bounded deviation from schedule)**.** *If* $F < 1$*, then* $\sigma_\varepsilon^2/\sigma^2 = \sum_{j=0}^{\infty} \sum_i (f_{i|j})^2 \leq 1/(1-F^2)$.

Proof. An upper bound of $\sum_{j=0}^{\infty} \sum_i (f_{i|j})^2$ is obtained by choosing the $|f_{i|j}|$ that maximizes $\sum_i (f_{i|j})^2$ subject to $\sum_i |f_{i|j}| \leq F^j$, as per the previous lemma. Since the objective function $\sum_i (f_{i|j})^2$ is convex, its maximum is reached at the vertices of the feasible region, see [63]. The vertices of the feasible region here are: either $|f_{i|j}| = 0$, $\forall i$; or $|f_{i|j}| = F^j$ for a single i and 0 for all the other i. It's easy to see that the maximum is reached in any of the latter cases and therefore $\sum_i (f_{i|j})^2$ is bounded from above by F^{2j}. Thus $\sum_{j=0}^{\infty} \sum_i (f_{i|j})^2 \leq \sum_{j=0}^{\infty} F^{2j} = 1/(1-F^2)$ if $F < 1$. □

Theorem 3.2.2 shows that any linear control method whose coefficients satisfy $F = \sum_i |f_i| < 1$ will adhere to a schedule with bounded deviations from it. Thus, such a control will be able to guarantee a certain degree of reliability.

On the other hand, linear headway-based control methods such as (3.7), (3.8) or (3.9) have $F \geq 1$; so theorem 3.2.2 does not apply. It is now shown (see Theorem 3.2.5) that all linear headway-based control methods with a finite number of non-zero coefficients have an unbounded measure of schedule reliability, $\sigma_\varepsilon^2(\boldsymbol{f})$. The methods in question are of the form:

$$D_{n,s} = d_{n,s} + \sum_i \alpha_i h_{n-i,s}, \tag{3.12}$$

which includes (3.7a), (3.8a), and (3.9a) as special cases.

Lemma 3.2.3. *Define* $F' = \sum_i f_i$*, then* $\sum_i f_{i|j} = (F')^j$*,* $\forall j \geq 1$*.*

Proof. Similarly as in lemma 3.2.1, let us expand the j^{th} convolution term: $\sum_i f_{i|j} = \sum_i \sum_k f_{i-k} f_{k|j-1} = \sum_k f_{k|j-1} \sum_i f_{i-k} = F' \sum_k f_{k|j-1} = F' \sum_i f_{i|j-1}$, $\forall j \geq 1$. These steps can be applied recursively getting: $\sum_i f_{i|j} = (F')^j$, $\forall j \geq 1$. □

Lemma 3.2.4 (Sufficient condition for unbounded deviation from a schedule)**.** *If* $\boldsymbol{f}$ *has at most* $2n+1$ *non-zero coefficients centered around* $i = 0$ *so that* $F' = \sum_i f_i = \sum_{i=-n}^{n} f_i = 1$ *and* $f_i = 0$ *for all other* i*, then* $\sigma_\varepsilon^2/\sigma^2 = \sum_{j=0}^{\infty} \sum_i (f_{i|j})^2$ *is unbounded.*

Proof. Notice that the index of the non-zero terms of the n^{th} convolution must be in the interval $[nj, nj]$. Thus $\sum_i (f_{i|j})^2 = \sum_{i=-nj}^{nj} (f_i \mid j)^2$, which consists of $2nj+1$ terms. A lower bound to $\sum_{i=-nj}^{nj} (f_{i|j})^2$ is obtained by choosing the $f_{i|j}$ values that minimize it, subject to $\sum_{i=-nj}^{nj} f_{i|j} = 1$ as per Lemma 3.2.3. The minimum arises when all the terms in the summation are equal, i.e., when $f_{i|j} = 1/(2nj+1)$, $\forall i \in [-nj, nj]$:

$$\sum_i (f_{i|j})^2 = \sum_{i=-nj}^{nj} (f_i \mid j)^2 \geq \frac{2nj+1}{(2nj+1)^2} = \frac{1}{2nj+1}.$$

Since the sum $\sum_{j=0}^{\infty} \frac{1}{2nj+1}$ diverges because it is a special case of the general harmonic series, so does $\sigma_{\varepsilon^2}/\sigma^2 = \sum_{j=0}^{\infty} \sum_i (f_{i|j})^2$ when $F' = 1$. □

Theorem 3.2.5 (Unbounded schedule deviations of headway-based linear control methods)**.** *All headway-based linear control methods with a finite number of non-zero coefficients, including (3.7), (3.8), and (3.9), exhibit unbounded deviations from the schedule,* $\sigma_\varepsilon^2(\boldsymbol{f}) = \infty$*.*

Proof. Since the number of non-zero coefficients is finite, the indices of these coefficients must be contained in an interval $[n, n]$ for a sufficiently large n (some of the coefficients in the interval may be zero). Thus, in view of Lemma 3.2.4, it suffices to show that $F' = 1$ for headway-based linear control methods. To see this, express (3.12) in terms of schedule deviations:

$$D_{n,s} = d_{n,s} + \sum_i \alpha_i h_{n-i,s} = (d_{n,s} + H \sum_i \alpha_i) + \sum_i (\alpha_i \varepsilon_{n-i,s} - \alpha_i \varepsilon_{n-i-1,s}).$$

Note from above that $\gamma_i = \alpha_i - \alpha_{i-1}$ and therefore $\sum_i \gamma_i = 0$ if the number of coefficients is finite. Then it follows from (3.4d) that $F' = \sum_i f_i = 1$. □

To sum up, any linear control method whose coefficients satisfy $F = \sum_i |f_i| < 1$ will exhibit bounded deviations from a schedule, while headway-based methods (which have

$F' = 1$) cannot provide this type of guarantee.[10] Bounded schedule deviations are essential in long headway services. In those cases, passenger arrivals are not uniform in time, but adjust to an advertised schedule. In, fact it has been shown multiple times that if passengers choose their arrival times to minimize their wait times, then their average waiting time is proportional to the buses average deviations from their schedule [16, 22].

For high frequency service schedule adherence is not important, only headway reliability. This is an easier goal to meet as the next subsection shows.

Headway Adherence

As in the previous subsection, let us first define the metric of interest, which in the present case is the function $\sigma_h^2(\boldsymbol{f}, n, s)$ that returns the headway variance $\text{var}(h_{n,s})$ given $\boldsymbol{f}$, n and s. We know that the headways $h_{n,s}$ can also be expressed as $h_{n,s} = H + \varepsilon_{n,s} - \varepsilon_{n-1,s}$ if there is a homogeneous planned headway H. Therefore, we see from (3.10b) that:

$$h_{n,s} = H + \sum_{j=0}^{s-1} \sum_{i} (f_{i|j} - f_{i-1|j}) \nu_{n-i,s-j}. \tag{3.13a}$$

Since the noise terms are i.i.d., $\text{var}(h_{n,s})$ is a sum of non-negative terms. As such, it is bounded above by the instance with $n, s \to \infty$:

$$\sigma_h^2(\boldsymbol{f}) \equiv \lim_{\substack{n\to\infty \\ s\to\infty}} \text{var}(h_{n,s}) = \sigma^2 \sum_{j=0}^{\infty} \sum_{i} (f_{i|j} - f_{i-1|j})^2 \geq \sigma_h^2(\boldsymbol{f}, n, s). \tag{3.13b}$$

Now, it is also possible to demonstrate as a corollary of Theorem 3.2.2 that all the control methods which satisfy the conditions in that theorem will also have bounded headway variances.

Corollary 3.2.6. *If $\sigma_\varepsilon^2(\boldsymbol{f})$ is bounded, then $\sigma_h^2(\boldsymbol{f})$ is also bounded.*

Proof. The headway variance upper bound $\sigma_h^2(\boldsymbol{f})$ satisfies

$$\begin{aligned} \sigma_h^2(\boldsymbol{f}) \equiv \lim_{\substack{n\to\infty \\ s\to\infty}} \text{var}(h_{n,s}) &= \lim_{\substack{n\to\infty \\ s\to\infty}} \text{var}(\varepsilon_{n,s} - \varepsilon_{n-1,s}) \\ &= \lim_{\substack{n\to\infty \\ s\to\infty}} [\text{var}(\varepsilon_{n,s}) + \text{var}(\varepsilon_{n-1,s}) + 2\text{cov}(\varepsilon_{n,s}, \varepsilon_{n-1,s})] \\ &\leq 4\sigma_\varepsilon^2(\boldsymbol{f}). \end{aligned}$$

[10] Note that these stability results could also be derived in a system where the noise parameters are not homogeneous across stops. In such case, the ratio $\sigma_\varepsilon^2/\sigma_{\max}^2$ would be bounded as in Theorem 3.2.2, where $\sigma_{\max}^2$ would simply be the maximum noise value that the buses would face in their motion: $\sigma_{\max}^2 = \max_{n,s} \sigma_{n,s}^2$.

The last inequality is true because $\text{cov}(\varepsilon_{n,s}, \varepsilon_{n-1,s}) \leq \sqrt{\text{var}(\varepsilon_{n,s})\text{var}(\varepsilon_{n-1,s})}$. Thus if $\sigma_\varepsilon^2(\boldsymbol{f})$ is bounded, $\sigma_h^2(\boldsymbol{f})$ is also bounded. □

In the case of headway-based control methods, both the forward-headway and two-way-headway methods (3.7) and (3.8) have been shown to have a bounded headway variance [23, 26]. On the other hand, numerical calculations of (3.13b) show that the backward-headway method (3.9) only produces bounded headway variances for low to medium demand levels ($\beta_{n,s} < 0.2$). Moreover, when $\beta_{n,s} < 0.2$ the control coefficient needs to be carefully chosen ($\alpha \approx 0.5$).

3.3 Optimal Control

This section presents a possible approach to select the control coefficients and the slack values. It is proposed to choose the control coefficients $\boldsymbol{f}$ that minimize the sum of the slack times $d_{n,s}$ required to avoid negative holding times with high probability while guaranteeing that the standard deviations from the schedule do not surpass a maximum value: i.e., a given level of schedule reliability. This proposal is reasonable in a homogeneous bus line because slack is inversely related to commercial speed.

To obtain slack times $d_{n,s}$ that avoid negative holding times with high probability, one can combine equations (3.4b) and (3.10), so that the holding time is expressed as a function of the control coefficients and noise terms:

$$\begin{aligned} D_{n,s} &= d_{n,s} - [(1+\beta_{n,s})\varepsilon_{n,s} - \beta_{n,s}\varepsilon_{n-1,s} - \sum_k f_k \varepsilon_{n-k,s}] \qquad (3.14a) \\ &= d_{n,s} - \sum_{j=0}^{s-1}\sum_i [(1+\beta_{n,s})f_{i|j} - \beta_{n,s}f_{i-1|j} - f_{i|j+1}]\nu_{n-i,s-j}. \end{aligned}$$

Under the assumption of i.i.d. noise, the variance of the holding time $D_{n,s}$ is the sum of the variances of many independent random variables, which as before is bounded above by the limiting case; i.e.:

$$\begin{aligned} \text{var}(D_{n,s}) \leq \sigma_D^2(\boldsymbol{f}, \beta_{n,s}) &\equiv \lim_{\substack{n\to\infty \\ s\to\infty}} \text{var}(D_{n,s}) \qquad (3.14b) \\ &= \sigma^2 \sum_{j=0}^{\infty}\sum_i [(1+\beta_{n,s})f_{i|j} - \beta_{n,s}f_{i-1|j} - f_{i|j+1}]^2. \end{aligned}$$

According to the central limit theorem, $D_{n,s}$ is approximately normal. Therefore, to ensure that the holding time is rarely negative, i.e., $\Pr\{D_{n,s} < 0\} \approx 0$, we shall choose

$$d_{n,s}(\boldsymbol{f}, \beta_{n,s}) = 3\sigma_D(\boldsymbol{f}, \beta_{n,s}), \qquad (3.14c)$$

so that the assumption is true 99.87% of the time.

The optimization problem for a line with J control points and N buses is then the following mathematical program, where s_ε is a guaranteed standard deviation from the schedule (chosen by the transit operator).

$$\text{(MP3.1)} \qquad \min_{\boldsymbol{f}} \sum_{s=0}^{J} \sum_{n=0}^{N} \beta_{n,s} d_{n,s}(\boldsymbol{f}, \beta_{n,s}) \quad \text{s.t. } \sigma_\varepsilon(\boldsymbol{f}) \leq s_\varepsilon$$

The functions in (MP3.1) are given by (3.11b), (3.14b) and (3.14c), and can be calculated numerically. A similar program can be written if the objective is to guarantee a headway standard deviation by using σ_h and (3.13b), instead of σ_ε and (3.11b) in the constraint.

3.3.1 Sensitivity Analysis: Homogeneous Line

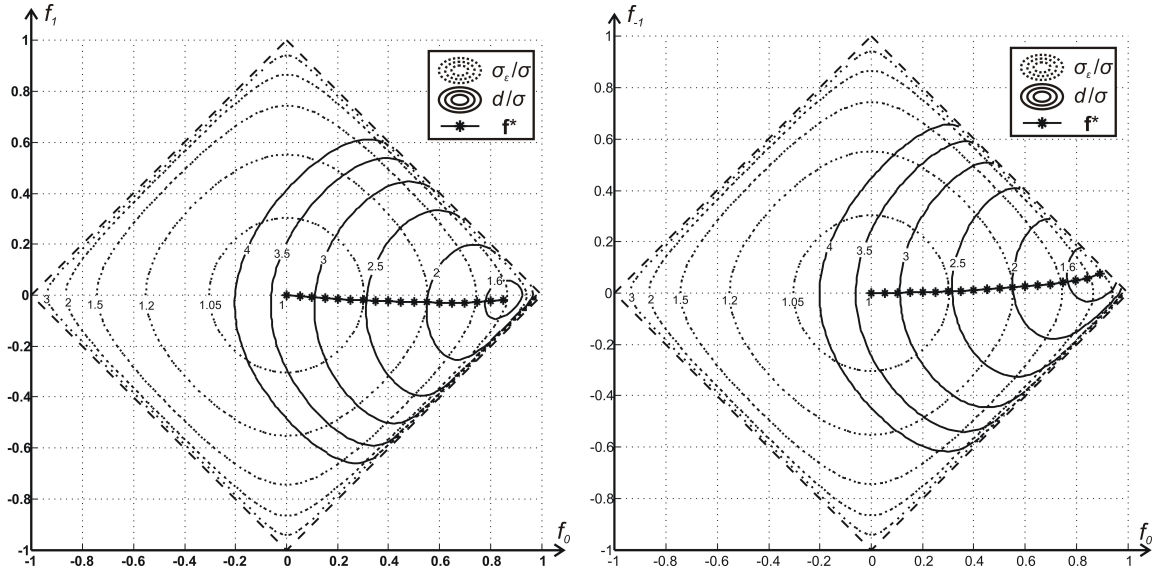
Note that in (MP3.1) the dimensionless demand rates, $\beta_{n,s}$ can be different at different stations (from 1 to J) and for different buses(from 1 to N); and so can the slacks, $d_{n,s}$. For simplicity of exposition it is assumed here that the demand rate is homogeneous in time and space ($\beta_{n,s} = \beta$). This is a worst case scenario.[11] It is of interest because if a strategy can prevent buses from bunching in the most challenging situation, it should be able to do so in other situations.

With uniform demand rates in time and space, the slack time is forced to be homogeneous across stations and buses ($d_{n,s} = d$, $\forall n, s$). Thus the subscripts n, s are now dropped, and (MP3.1) becomes:

$$\text{(MP3.2)} \qquad \min_{\boldsymbol{f}} d(\boldsymbol{f}, \beta) \quad \text{s.t. } \sigma_\varepsilon(\boldsymbol{f}) \leq s_\varepsilon$$

Figure 3.1 shows the contour lines of $\sigma_\varepsilon(\boldsymbol{f})/\sigma$ and $d(\boldsymbol{f})/\sigma$, when $\beta = 0.1$ for two methods that have two non-zero control coefficients. In Figure 3.1a, all $f_i = 0$ except for f_0 and f_1, and in Figure 3.1b, only f_0 and f_{-1} are non-zero. The interior of the dashed squares in the figures are the regions where the condition $F = \sum_i |f_i| < 1$ holds so that stability is guaranteed. We see that in either case, both $\sigma_\varepsilon(\boldsymbol{f})$ and $d(\boldsymbol{f})$ are quasi-convex functions of

[11] To illustrate this, let us focus on the motion of a single bus and assume that the total demand is fixed, the trip time is deterministic, and all buses are initially on time. In this case, the motion equation in (3.4c) becomes $\varepsilon_{n,s+1} = (1 + \beta_{n,s})\varepsilon_{n,s}$. Now insert a noise term to bus n between stations 0 and 1 so that $\varepsilon_{n,1} = \nu_{n,1}$, and see how the noise propagates. It follows from the initial motion equation that $\varepsilon_{n,s+1} = \nu_{n,1} \prod_{i=1}^{s}(1 + \beta_{n,i})$. Given that the total demand for that particular bus $\sum_{i=1}^{s} \beta_{n,i}$ is fixed, $\varepsilon_{n,s+1}$ is maximized when all $\beta_{n,i}$'s are the same, i.e., when demand is uniform across all stations for that particular bus.



(a) All $f_i = 0$ except for f_0 and f_1.

(b) All $f_i = 0$ except for f_0 and f_{-1}.

Figure 3.1: The iso-$\sigma_\varepsilon/\sigma$ and iso-d/σ contours and optimal control $\boldsymbol{f}*$ with two coefficients and $\beta = 0.1$. Dotted lines within the square are the contours with equal $\sigma_\varepsilon/\sigma$-value and solid lines are the contours with equal d/σ-value. The dashed square limits the stability region for schedule deviations.

$\boldsymbol{f}$ within this stability region. Clearly the optimal control coefficient values for any s_ε are at the point where its $\sigma_\varepsilon = s_\varepsilon$ contour is tangent to a d-contour, with the two gradients pointing against each other. This geometrical argument shows that the reliability constraint is always binding. Figure 3.1 shows the loci of optimal control coefficients for different $\sigma_\varepsilon/\sigma$ levels by means of dark diamonds.

Since contours are convex, the solutions were obtained with a local gradient search. See Appendix A for the derivation of the gradients. This method works well with up to 11 non-zero control coefficients, which we have tested. Note from Figure 3.1 that the schedule-based control method (with $f_i = 0, \forall i$) is actually among the optimal solutions. It provides the best possible schedule reliability (always departing on time), though it requires very high levels of slack ($d/\sigma \approx 3.4$).

Note also that the optimal control coefficients in both cases are very close to the f_0 axis. This indicates that the optimal f_1^* and f_{-1}^* are very small, and that the performance of a control method with a single non-zero coefficient ($f_0 \neq 0$) may be comparable with that with two or more non-zero coefficients. Table 3.1 confirms this guess. It shows the optimal slack time (in units of σ) for $\beta = 0.1$. Different demand rates yield similar results.

Figure 3.2 shows the optimal control coefficients f_{-2}^*, f_{-1}^*, f_0^*, f_1^*, and f_2^* for different β and s_ε/σ. Note that the optimal control coefficients are fairly insensitive to changes in the dimensionless demand rate, β. This is good news, because in reality we may not know the

Table 3.1: Effect of the number of non-zero coefficients on d^*/σ when $\beta = 0.1$.

	$s_\varepsilon/\sigma = 1$	$s_\varepsilon/\sigma = 1.2$	$s_\varepsilon/\sigma = 1.5$	$s_\varepsilon/\sigma = 2$
$f_i = 0\,\forall i$, except f_0	3.314	1.989	1.657	1.527
$f_i = 0\,\forall i$, except f_{-1}, f_0, f_1	3.314	1.978	1.637	1.463
$f_i = 0\,\forall i$, except $f_{-2}, f_{-1}, f_0, f_1, f_2$	3.314	1.978	1.637	1.463

demand rate very well. Also note that the values of f^*_{-2}, f^*_{-1}, f^*_1 and f^*_2 are roughly negligible (absolute values less than 0.05 in Figure 3.2). The other optimal control coefficients are not shown in the figure because they are even smaller. So, the performance of a control method with only one non-zero coefficient ($f_0 \neq 0$), which will be named "simple control", should be near-optimal. This seems to be the case. Numerical calculations show that for $\beta \in (0.01, 0.1)$ and $s_\varepsilon/\sigma \in (1, 2)$, the slack time with the simple control is always within 4% of that of the optimal control obtained with 11 control coefficients.

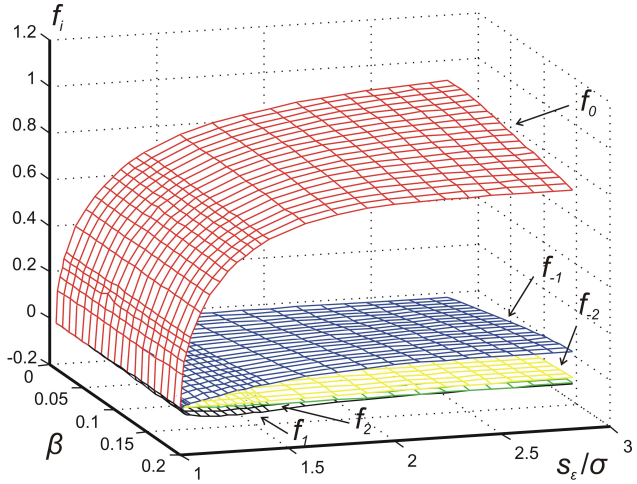
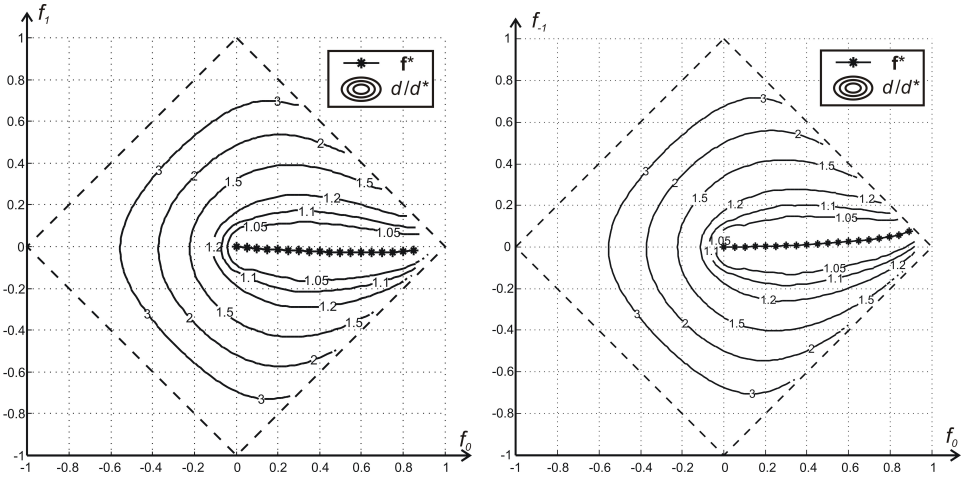


Figure 3.2: Optimal values of f^*_{-2}, f^*_{-1}, f^*_0, f^*_1, and f^*_2 with different demand rates and schedule reliability values.

The strategy is also robust to the selection of the control coefficients. Figure 3.3 shows that control coefficients that do not differ much from the optimal control coefficients do not increase the required slack time much. The figure displays the ratio of d/d^* when $\beta = 0.1$ and the general control method has two non-zero coefficients (f_0 and f_1 or f_{-1} and f_0). In this figure, d is the slack time of the non-optimal control coefficients at the indicated point and d^* is the optimal slack time for the σ_ε value of the point. We see that within a small neighborhood around the optimal coefficients $\boldsymbol{f}^*$, the ratio is close to 1 but large deviations in the selection of the control coefficients can result in inefficiency.

The near-optimality of the simple control is nice for both implementation and theoretical analysis. From the implementation point of view, only the arrival times of the current bus and its leading bus, as well as the virtual schedule, are needed to decide the holding time



(a) All $f_i = 0$ except for f_0 and f_1. (b) All $f_i = 0$ except for f_0 and f_{-1}.

Figure 3.3: Slack sensitivity to the choice of control coefficients when $\beta = 0.1$. Each point on the plane is associated with a reliability level σ_ε. The contour value at a point is the ratio of the non-optimal slack time d associated with that point over the optimal slack time d^* corresponding to the σ_ε value of the point.

of the current bus at a given station. From an analysis point of view, the simple control method is helpful because formulas simplify and (MP3.1) can be solved in closed form. These simplifications are described below.

3.3.2 A Simple Control Method

The control law, bus motion, and metrics of interest are re-derived below, with f_0 as the only non-zero control coefficient, and therefore the only decision variable. For the system to be stable: $F = |f_0| < 1$. Note that $f_{i|j} = (f_0)^j \delta(i)$, where $\delta(i)$ is the discrete unit impulse function. In this case, equations (3.4b), (3.4c), (3.11b), (3.13b), (3.14b), and (3.14c) reduce to:

$$D_{n,s} = d_{n,s} - [(1 + \beta_{n,s} - f_0)\varepsilon_{n,s} - \beta_{n,s}\varepsilon_{n-1,s}], \tag{3.15a}$$

$$\varepsilon_{n,s+1} = f_0\varepsilon_{n,s} + \nu_{n,s+1}, \tag{3.15b}$$

$$\sigma_\varepsilon^2(f_0) = \sigma^2/(1 - f_0^2), \tag{3.15c}$$

$$\sigma_h^2(f_0) = 2\sigma^2/(1-f_0^2), \tag{3.15d}$$

$$\sigma_{D_{n,s}}^2(f_0, \beta_{n,s}) = \frac{\sigma^2[(1+\beta_{n,s}-f_0)^2+\beta_{n,s}^2]}{1-f_0^2}, \tag{3.15e}$$

and

$$d(f_0, \beta) = 3\sigma_{D_{n,s}}(f_0, \beta_{n,s}). \tag{3.15f}$$

Note that these equations only apply if the system is stable (i.e., if $f_0^2 < 1$); then $\sigma_\varepsilon^2(f_0) > \sigma^2$.

The optimal solution of (MP3.1) is obtained by choosing an $f_0 \in [0,1)$ that minimizes the sum of (3.15e) over n and s such that (3.15c) is bounded above by s_ε^2, where $s_\varepsilon > \sigma$. Note that (3.15e) is a convex function of f_0 with a minimum at $f_0^{\min} \equiv \frac{1+\beta_{n,s}+\beta_{n,s}^2-\beta_{n,s}\sqrt{\beta_{n,s}^2+2\beta_{n,s}+2}}{1+\beta_{n,s}}$. Moreover, (3.15c) is strictly increasing with respect to $f_0 \in [0,1)$. Thus, in the optimal solution, the constraint will be binding as long as $s_\varepsilon^2 \leq \sigma_\varepsilon^2(f_0^{\min})$. In that case, the optimal coefficient for the simple control can be obtained using (3.15c) since s_ε^2 will match the actual variance σ_ε^2. On the other hand, when $s_\varepsilon^2 > \sigma_\varepsilon^2(f_0^{\min})$ the optimal coefficient will be $f_0^* = f_0^{\min}$. Therefore, the optimal coefficient for the simple control method can be expressed as:

$$f_0^* = \min\{\frac{1+\beta_{n,s}+\beta_{n,s}^2-\beta_{n,s}\sqrt{\beta_{n,s}^2+2\beta_{n,s}+2}}{1+\beta_{n,s}}, \sqrt{1-(\sigma^2/s_\varepsilon^2)}\}, \tag{3.15g}$$

and the optimal slack is

$$d_{n,s}^* = 3s_\varepsilon\sqrt{(1+\beta_{n,s}-\sqrt{1-(\sigma^2/s_\varepsilon^2)})^2+\beta^2}, \tag{3.15h}$$

if $f_0^* = \sqrt{1-(\sigma^2/s_\varepsilon^2)}$.

3.3.3 Control Method Comparison: Balancing Headway Regularity and Slack Time

The initial proposed optimization framework consisted of the minimization of slack time (to maximize commercial speed) subject to a schedule reliability constraint (or similarly a headway regularity constraint). This formulation is appealing because it does not require knowledge of the complete passenger origin-destination table. For simplicity, here it is assumed that only the average user trip length l needs to be known. With l known, one can

balance the two metrics by minimizing the sum of the average passenger waiting and riding times.

Here we focus on bus lines operating with short headways and compare the simple control to the other four methods (schedule-based, forward headway, backward headway and two-way headway) using a headway regularity metric. For bus lines with long headways, headway-based control methods are not applicable. Thus, the simple control method would compare even better.

It is assumed for the comparisons that: (i) the average bus cruising speed is v_c (i.e., including acceleration and deceleration due to the stops); (ii) the station spacing X is uniform; (iii) the unitary demand β is uniform; (iv) passengers value their waiting time γ times ($\gamma > 1$) more than their riding time.

Under these conditions, the average passenger waiting time is $H/2 + \sigma_h^2/(2H)$. The passenger riding time is $l/v_c + (\beta H + d)l/X$. The weighted sum of the two can be expressed as the sum of a fixed component T_0 and a component ΔT that depends on the control method:

$$T_0 = \gamma \frac{H}{2} + \frac{l}{v_c} + \frac{\beta H l}{X}, \tag{3.16a}$$

$$\Delta T = \gamma \frac{\sigma_h^2}{2H} + \frac{dl}{s}. \tag{3.16b}$$

Note T_0 includes waiting, line-haul riding and loading time under ideal conditions. It is the minimal feasible trip time, and it is achievable only if the bus system has no disturbances. The added term ΔT includes the extra time penalty passengers suffer due to non-uniform headways (σ_h^2) and to the slack time (d). The variables σ_h^2 and d can be calculated as a function of the control coefficients with equations (3.13b), (3.14b), and (3.14c) for all considered methods. For each method the control coefficients that minimize (3.16b) are obtained, and the optimum ΔT is calculated.

Figure 3.4 shows the optimum ratio $\Delta T/T_0$ for all the control methods, as a function of the demand rate β and the dimensionless trip length l/X when the control is applied at all stations. The following parameter values were used: $X = 400$ meters, $H = 5$ minutes, $v_c = 20$ km/hr, $\gamma = 2$ and $\sigma = 10$ seconds. The figure shows that for low demand rates ($0.01 < \beta < 0.05$) all the headway-based methods, and that the simple control perform much better than the schedule-based method. The $\Delta T/T_0$ ratios in this situation with low demand are in the range of 4-11% for the backward headway method, 5-10% for the forward headway method and only 3-7% for the two-way headway method and the simple control. As demand increases, the only qualitative difference is that the backward headway method becomes unstable. Note that the two-way headway method and the simple method are practically indistinguishable, with a maximal $\Delta T/T_0$ ratio of 15% at $\beta = 0.2$ and $l/X = 25$. It should be said in fairness that the results of the two-way headway method and the backward headway

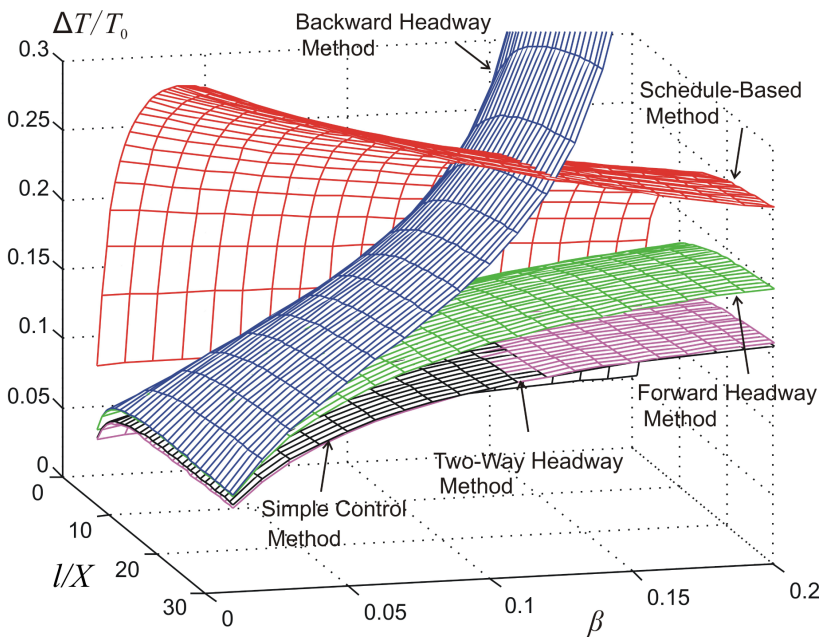


Figure 3.4: Ratio of variable trip time over fixed trip time ($\Delta T/T_0$) for different control methods with various demand rates β and dimensionless trip lengths l/X, with control points located at each station.

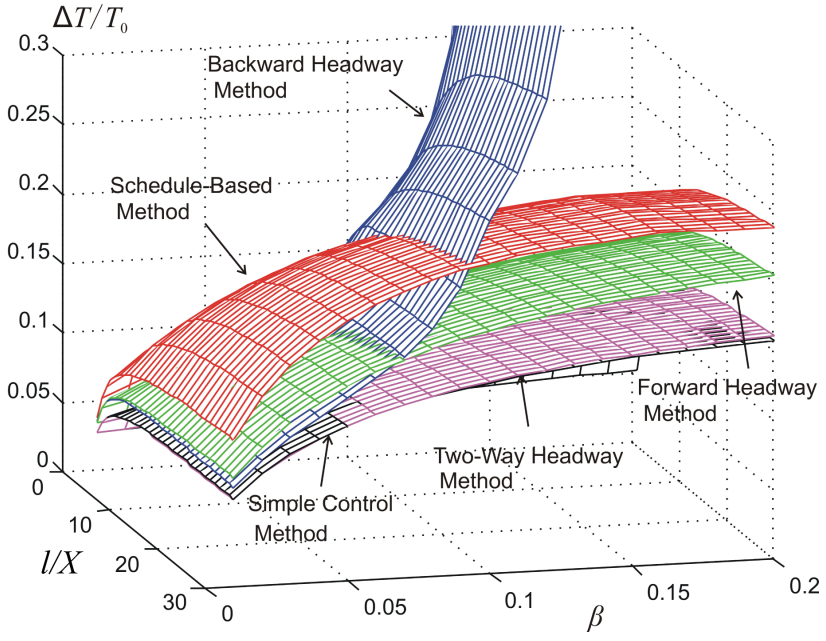


Figure 3.5: Ratio of variable trip time over fixed trip time ($\Delta T/T_0$) for different control methods with various demand rates β and dimensionless trip lengths l/X, with control points located optimally.

method should be slightly optimistic because in reality the backward headway information will not be readily available and needs to be estimated.

Figure 3.5 shows the same information as Figure 3.4, assuming that the control points are also optimally located (every K^* stations instead than at every station) as per Appendix B. The value of K^* is different for different methods and different combinations of β and l/X. In the minimization of (3.16b), the headway variance σ_h^2 is now replaced by its average across all stations $\bar{\sigma}_h^2$, calculated as explained in the last paragraph of Appendix B. By allowing values $K > 1$, the slack time needed for the schedule-based method is greatly reduced for low demand as shown in the figure, because in this case $K^* >> 1$. The backward headway method also saves a little slack time because in this case $K^* \approx 2$. The results for the simple control method, the forward headway method and the two-way headway method remain the same as in Figure 3.4, because for these methods $K^* = 1$.

Note that the simple method reduces slack time by nearly 40% compared with the schedule-based method. The two-way headway method behaves similarly. For moderately heavy demand ($\beta = 0.1$), these two methods reduce the slack time of the schedule-based method by more than 50%, and reduce the commercial speed compared with the ideal by only about 10%. These predictions are optimistic, however, for two-way method because they assume that backward headways can be predicted accurately. Furthermore the two-way method cannot help the buses keep to a schedule. Thus, it can be concluded that the simple control outperforms all other methods whether or not buses follow a schedule.

Chapter 4

Single Line Case Study: UC Berkeley Campus Shuttle

The previous chapter presented a general framework for linear dynamic holding control methods in isolated lines. The scope was mainly theoretical and it focused on deriving general insights. The chapter also analyzed the trade-off that emerges between guaranteeing a certain reliability level and the commercial speed of the buses in the line, but in a highly idealized scenario.

This chapter takes a more realistic approach and it illustrates how one could tune the control coefficients balancing the reliability and commercial speed of a real bus line. The optimization problem that leads to the selection of the slack times and control coefficients is adapted to a non-homogeneous system. This chapter also introduces a simulation tool that can compare the multiple dynamic holding methods presented earlier in realistic contexts.

4.1 System Description

Data for this analysis were obtained from an empirical study [5] in which the author participated. This work examined the operating conditions in the UC Berkeley campus shuttle system (known as Bear Transit) in April 2011. During that time, the Bear Transit system included three routes: 1) the Perimeter route, which ran clockwise around campus; 2) the Reverse Perimeter route, which ran counter-clockwise around campus; and, 3) the Central route, which bisected the campus in its longitudinal direction. A fourth route, the Hill route, was also in operation; however, this route was excluded from the study because it primarily served off-campus origins and destinations. The three routes in the study are displayed in Figure 4.1. The Perimeter line was served by 2 buses and had average headways of 12 minutes, while the remaining two lines were each served by a single bus with 30 and 20 minutes cycle times on the Reverse Perimeter and the Central route respectively.

The collected data included average demand rates, origin/destination patterns and travel time parameters. Two distinct tools were used to collect the data: (i) a passenger survey, and

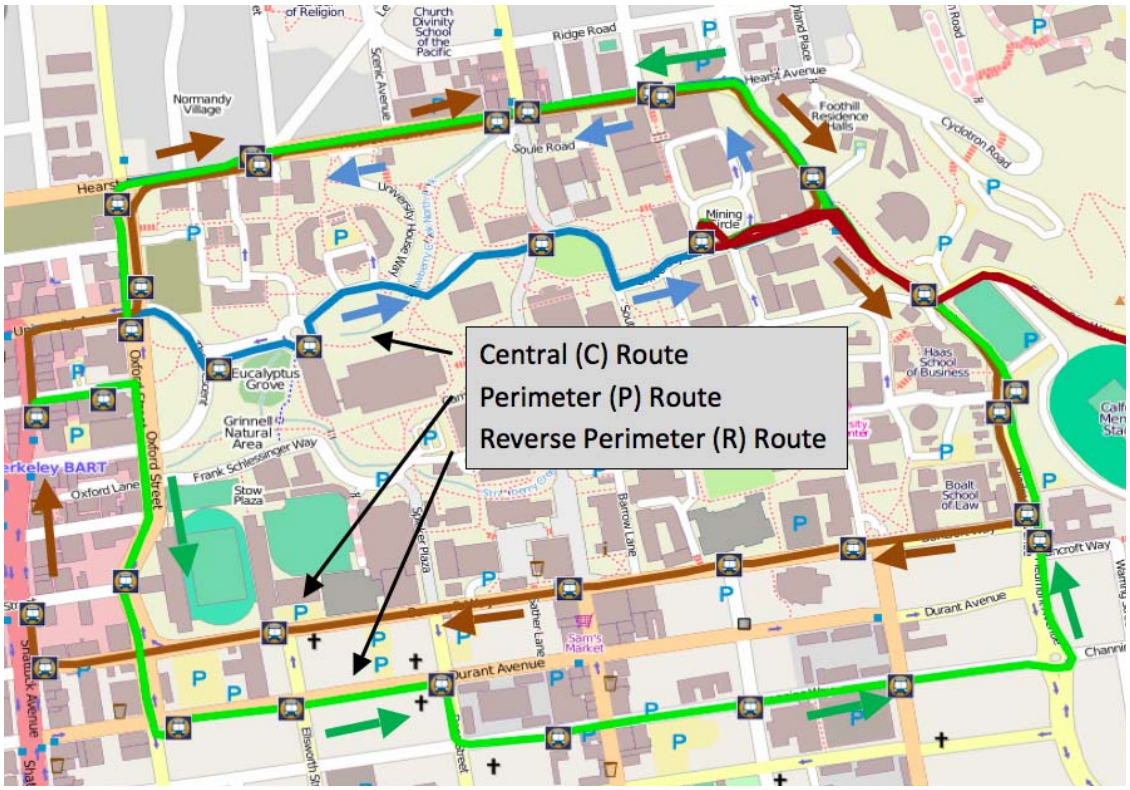


Figure 4.1: Routes operated by Bear Transit in the main UC Berkeley campus circa 2011.

(ii) a semi-automatic software application that ran on laptops to audit semi-automatically the operating conditions. The survey was given to all passengers that entered the buses during various times of the day and it was completed by approximately 95% of riders. It consisted of several questions that collected information on trip motivation, origins and destinations, and demographics. Responses revealed travel demand patterns on the Bear Transit System and highlighted key movements across the campus. The main result for the purpose of this dissertation was the estimation of an origin/destination matrix, which will be used later in this chapter.

The audit was performed by members of the research team who rode the system and collected information on all passenger movements (e.g., boardings and alightings) at every station and travel times between stations. Care was taken to ensure that the audit covered most operating hours on typical class days (Monday–Thursday) to get an accurate depiction of system use. See Appendix C for more information on the data collection procedures.

From this comprehensive examination, it was clear that the Perimeter route carried the vast majority of passengers (over 80%). By combining the data collected from the audit and survey, the study also revealed that average passenger travel times (including access, riding and waiting time) could be reduced by over 15% if the Reverse and Central routes were eliminated and their buses were moved to serve the Perimeter line, as depicted in

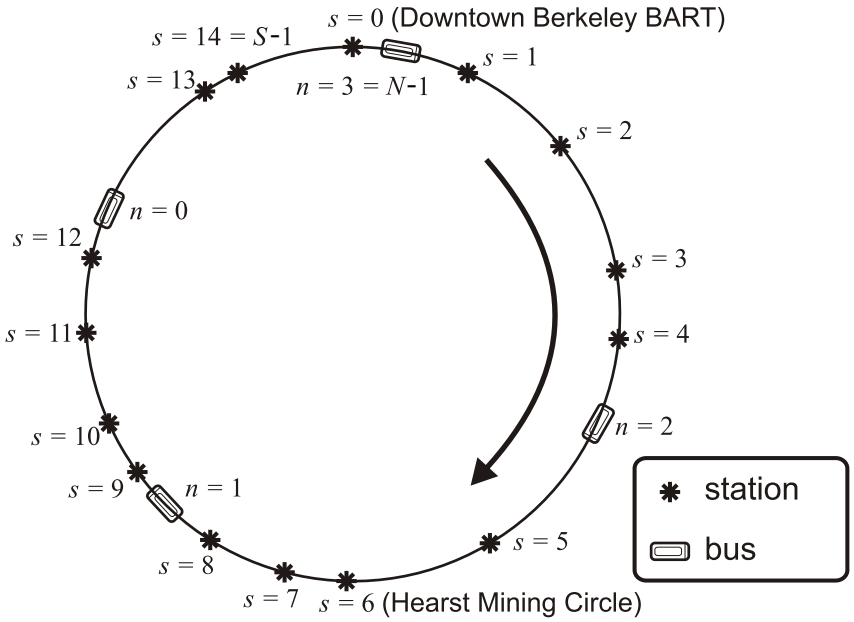


Figure 4.2: Graphical representation of the Perimeter bus line with $S = 15$ stations and $N = 4$ buses.

Figure 4.2. Unfortunately, such configuration could be problematic. Consolidating 4 buses on the perimeter route would result in smaller headways (about 6 minutes) and an increased need for good headway regularity.

This need motivated a performance analysis of the multiple linear holding control methods on the consolidated Perimeter line. Sections 4.2 and 4.3 describe the technical details

Moreover, this analysis also tries to cover a gap in the bus control literature, since no research had assessed the performance of the multiple linear holding methods described in Chapter 3 (presented in Chapter 3) under a common framework. Using a framework like the Bear Transit system is appealing because it provides the tools to derive a direct, meaningful, and realistic comparison of the different control methods.

4.2 Simulation Framework

To analyze how these control methods can affect system performance, an event-based simulation tool was created. The simulation models the operation of a system as a chronological sequence of events. In the case of a bus line, the simplest event set includes only three events: a bus arrival to a station, the end of the passengers' boarding and alighting, and the end of the holding time (if any).

Instead of simulating the evolution of the entire system state at regular time steps, this

approach simply keeps track of the most recent event. The simulation progresses through time by stepping through the different events chronologically. It achieves computational savings because not all time intervals need to be simulated. Figure 4.3 shows an idealized time-space diagram created using the event-based simulation tool for the case of four buses operating in a line. Note that the exact location of each vehicle is only known at its arrival to a station, when it completes passenger boarding and at the end of its holding. The location of the vehicles at all other times is inferred by linear interpolation.

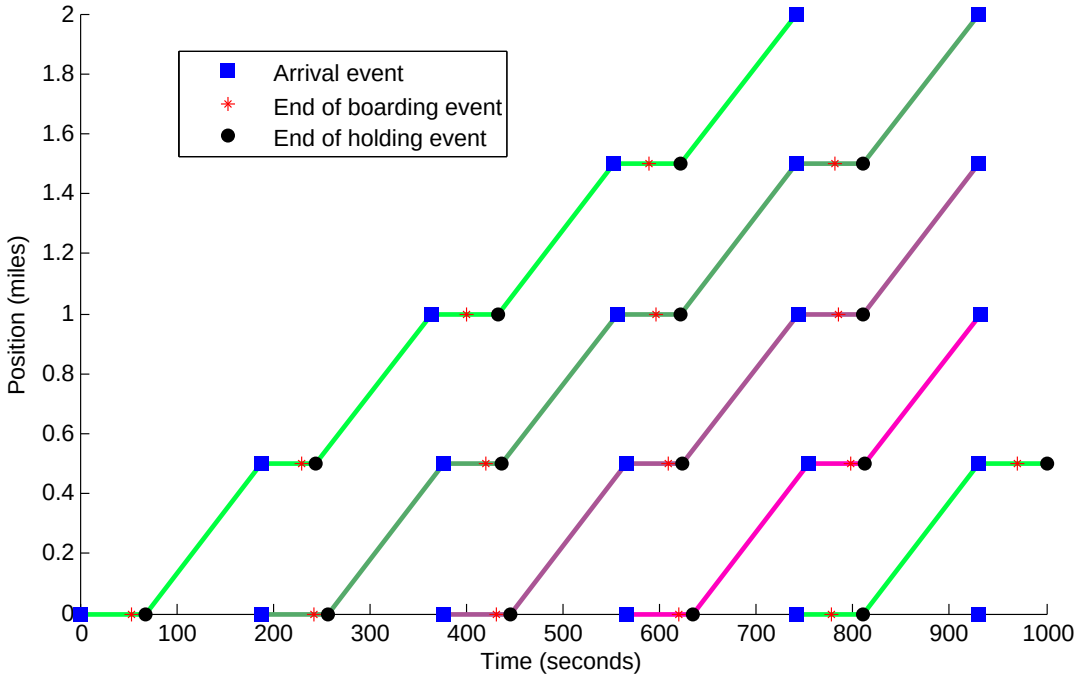
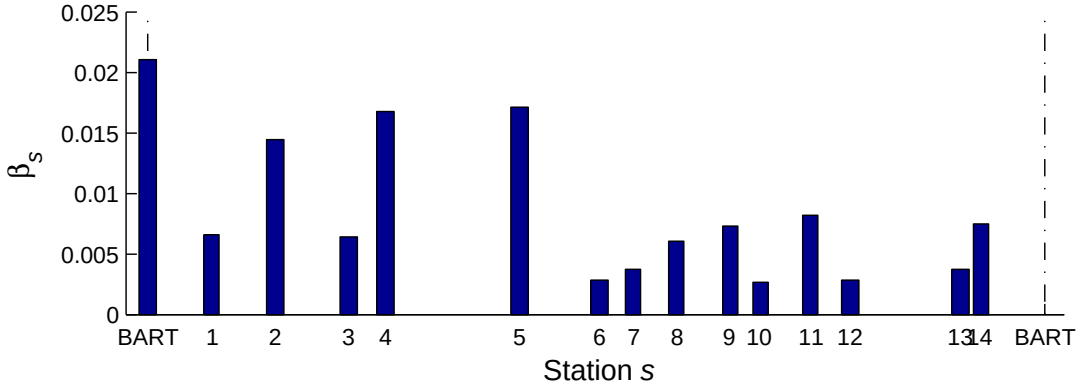


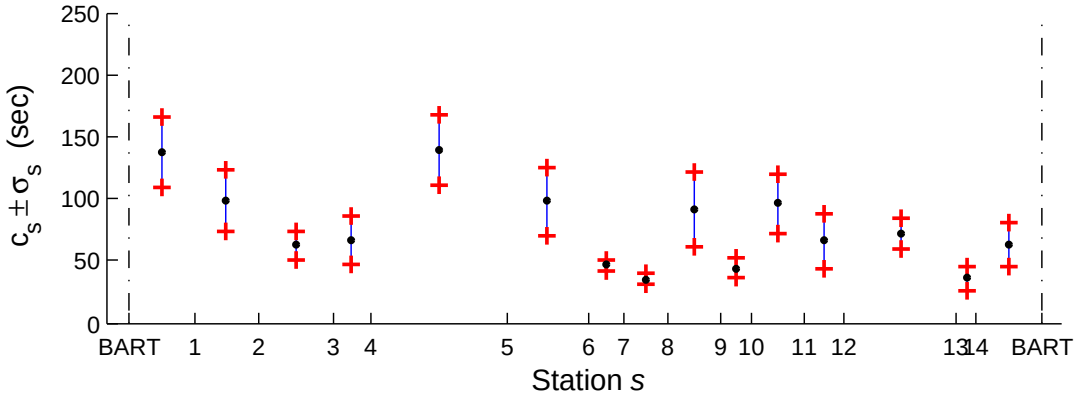
Figure 4.3: Time-space diagram reflecting the event-based simulation approach.

The key component in an event-based simulation is the logic behind the different event relations. For example, while the cruising time of a bus can be independent of any other event, the holding time of that same bus at a particular stop could depend on the arrival time of the previous bus at that same stop. The event relations in this simulation framework largely follow the assumptions introduced in Section 3.1.1. However, some of those assumptions are idealizations that were considered in order to obtain a theoretical model that is analytically tractable. In the simulation framework, some of these assumptions can be relaxed, yielding a more realistic representation of a bus line operation.

The first difference between the simulation and the theoretical framework is the fact that buses are continuously circulating in a closed loop rather than traveling along an infinite open line (the number of buses $N = 4$ and stops $S = 15$ are finite in the simulation). Note that this difference requires modifying the formulation in Section 3.1.2 to one with periodic boundary conditions. Any addition or subtraction of both the bus and station indices will



(a) Unitary demand parameter at the stops, β_s, sample obtained during weekdays between 10-12am (aggregated over the multiple Bear Transit lines using the O/D matrix).



(b) Mean $\pm$ standard deviation of the route links' travel times ($c_s \pm \sigma_s$), sample obtained during weekdays between 10-12am.

Figure 4.4: Demand and travel time data used in the consolidated Perimeter line simulations.

now be denoted using the symbols $\oplus$ and $\ominus$ respectively, as in [26]. These symbols indicate the addition or subtraction modulo N or S, depending on the index being operated on. For example, $n \oplus 1$ would indicate the bus following bus n, which in the case of $n = N - 1$ it would be bus 0. On the other hand, $s \ominus 1$ would indicate the station preceding station s, which in the case of $s = 0$ it would be station $S - 1$.

The simulation framework also uses real station spacings and travel time statistics collected from the field and shown in Figure 4.4b. In addition, for the sake of simplicity, the simulation uses the demand rates, shown in 4.4a, and origin/destination patterns obtained in the original Perimeter line. Thus, all these parameters are now station dependent, in contrast with the analysis in Sections 3.3.1, 3.3.2, and 3.3.3.

Another significant difference between the simulation and the theoretical analysis is the computation of the holding time. The assumption that the slack time is long enough so that holding time can almost never be negative allows for a linear formulation of the bus motion,

which makes an analytical solution possible. In reality, under extremely rare circumstances, the holding time originally given by equation (3.4b) may be a negative value and be physically impossible to apply.[1] The simulation overcomes the issue of negative holding times by using the following nonlinear truncation of (3.4b):

$$D_{n,s} = \max\{0, d_s + [(1+\beta_s)\varepsilon_{n,s} - \beta_s\varepsilon_{n\ominus 1,s}] + \sum_i f_i \varepsilon_{n\ominus 1,s}\}. \tag{4.1}$$

The simulation is also different because it considers a stochastic boarding process, where the number of passengers boarding at a station are obtained by generating a draw from a Poison process that considers both the demand rate and the actual inter-bus time. Finally, some of the holding methods involving backward headways, $h_{n\oplus 1,s} = \varepsilon_{n\oplus 1,s} - \varepsilon_{n,s}$, are handled more realistically. Backward headways are generally non-causal, because when bus n arrives at station s, bus $n \oplus 1$ has not arrived yet at that station. While the headway can be inferred from the buses spacing, in our simulation, we make the assumption that the schedule deviation of bus $n \oplus 1$ at station s equals its last known schedule deviation at the time when the holding for bus n, $D_{n,s}$, is calculated. As in the analysis of Chapter 3 the cruising times between stations are also stochastic. They are simulated by drawing random values from a lognormal probability distribution. The distributions were fitted to the field data using maximum likelihood estimation [21].

The simulation considers the operation of the Perimeter line for a 2 hour period. The simulation parameters were obtained based on field measurements collected from 10am to 12pm during Monday to Thursday for the span of a month. During that two-hour period, the demand and noise parameters are considered to be time independent but station dependent (i.e. $\beta_n, s = \beta_s$, $\forall n \in \{0, 1, \cdots, N-1\}$).

4.3 Control Optimization

The control parameters are optimized with the approach in Section 3.3.3. The goal once again is balancing reliability and commercial speed. In this case though, the different operational characteristics (looping route with a limited number of stops and buses) and the heterogeneity of demand, travel times, and noise, require modifying the formulation of Chapter 3. For instance, instead of considering an exogenous expected headway, the looping conditions of the line force us to treat the expected headway endogenously, depending on the number of buses and choice of slack times. The relationships that tie these variables

[1] Note that in the multiple simulation runs later performed, the likelihood of negative holding times as given by (3.4b) was very low (less than 1% of all bus holdings), when the slack time was calculated based on an equivalent equation to (3.14b) with a limited number of summation terms ($i \in \{0, 1, \cdots, S-1\}$) and non-homogeneous noise terms.

together are:

$$NH = \sum_{s=0}^{S-1} \beta_s H + c_s + d_s(\boldsymbol{f}, \beta_s), \tag{4.2a}$$

$$H(\boldsymbol{f}) = \frac{\sum_{s=0}^{S-1}(d_s(\boldsymbol{f}, \beta_s) + c_s)}{N - \sum_{s=0}^{S-1} \beta_s}. \tag{4.2b}$$

The bus motion equations will also differ slightly from the ones found in Chapter 3. They now are:

$$\varepsilon_{n,s\oplus 1} = \varepsilon_{n,s} + \beta_{n,s}(\varepsilon_{n,s} - \varepsilon_{n\ominus 1,s}) + (D_{n,s} - d_{n,s}) + \nu_{n,s\oplus 1}, \tag{4.2c}$$

$$D_{n,s} = d_{n,s} - [(1 + \beta_{n,s})\varepsilon_{n,s} - \beta_{n,s}\varepsilon_{n\ominus 1,s}] + \sum_i f_i \varepsilon_{n\ominus i,s}. \tag{4.2d}$$

The previous two equations can be combined obtaining the linear equation with white noise that characterizes the motion of controlled buses:

$$\varepsilon_{n,s\oplus 1} = \sum_i f_i \varepsilon_{n\ominus i,s} + \nu_{n,s\oplus 1}. \tag{4.2e}$$

Also, the steady state reliability parameters (i.e. variance of the deviations from schedule and headways) will now be station dependent, due to the heterogeneity of the noise terms:

$$\sigma^2_{\varepsilon,s}(\boldsymbol{f}) = \sum_{j=0}^{\infty} \sigma^2_{s\ominus j} \sum_i (f_{i|j})^2, \tag{4.2f}$$

$$\sigma^2_{h,s}(\boldsymbol{f}) = \sum_{j=0}^{\infty} \sigma^2_{s\ominus j} \sum_i (f_{i|j} - f_{i-1|j})^2, \tag{4.2g}$$

and the slack times will also be different between stations, since the variance of the holding times will now also be station dependent.

$$\sigma^2_{D,s}(\boldsymbol{f}, \beta_s) = \sum_{j=0}^{\infty} \sigma^2_{s\ominus j} \sum_i [(1 + \beta_s) f_{i|j} - \beta_s f_{i-1|j} - f_{i|j+1}]^2. \tag{4.2h}$$

Note that in principle, calculating the different steady state variances given by equations (4.2f), (4.2g), and (4.2h) requires the summation index j to be iterated ad infinitum. In practice however, those values generally converge to a finite number if the control coefficients $\boldsymbol{f}$ are appropriately chosen (e.g. $F < 1$ to guarantee bounded variance of the schedule deviations, see Section 3.2). Furthermore, in the case of the holding time variances (4.2h), numerical calculations using the consolidated Perimeter line data have shown that considering $j \in \{0, 1, 2, \cdots, 10S-1\}$, where $S = 15$, was sufficient enough to converge to the steady state bound within a 1% error, for any of the analyzed control methods when $F < 1$.

On the other hand, if the equations do not converge to a finite number (e.g. $F \geq 1$), the iteration can be truncated considering a reasonable iteration index limit. This would result in a proxy value for the worst case conditions in the line. In this particular case, the iteration limit was set equal to $10S - 1$, as explained in the previous paragraph. This is equivalent to considering that the buses in the line will operate for 10 consecutive cycles (exceeding the amount of cycles simulated in this analysis). In practice, this limit can be set by simply considering the total number of stops visited during the planned time for the daily operation of the buses.

A possible approach to optimize the selection of the control coefficients would be to minimize the average travel time under steady state conditions. This value can be obtained using the previous equations to calculate the variances of both headways and holding times. It also needs new parameters, which can be measured in the field, such as the demand rates at each stop, λ_s[2], as well as the origin/destination matrix, OD. Each of the elements of the matrix, $OD_{i,j}$, represent the probability that a rider boarding a bus at station i alights at station j. Based on these parameters one can compute the expected number of boardings at stop s, b_s, and the buses' occupancy at stop s once passenger boarding and alighting is completed, o_s, as:

$$b_s(\boldsymbol{f}) = H(\boldsymbol{f})\lambda_s, \tag{4.3a}$$

$$o_s(\boldsymbol{f}) = \sum_{i=0}^{S-1} b_{s\ominus i} \sum_{j=1}^{S-1-i} OD_{s\ominus i, s\oplus j}. \tag{4.3b}$$

The minimization problem to determine the choice of optimal control coefficients can then be formulated as:

[2]Note that λ_s can be expressed as the product between the unitless demand β_s and the average boarding time per passenger, t_b. In the Perimeter route, this parameter was estimated to be 2.7 seconds using linear regression.

$$\text{(MP4.1)} \qquad \min_{\boldsymbol{f}} \frac{\sum_{s=0}^{S-1}[(\gamma b_s(\boldsymbol{f})\frac{H(\boldsymbol{f})}{2}(1+\frac{\sigma_{h,s}^2}{H^2(\boldsymbol{f})}))+(o_s(\boldsymbol{f})d_s(\boldsymbol{f}))]}{H(\boldsymbol{f})\sum_{s=0}^{S}\lambda_s}, \quad \text{s.t. } d_s(\boldsymbol{f}) = 3\sigma_{D,s}(\boldsymbol{f},\beta_s).$$

MP4.1 assumes that the headway is short enough for a schedule not to be published, so passengers arrive randomly at the station. The objective function ignores all travel time terms that do not depend on the choice of the control coefficients. It only accounts for the expected wait time that passengers will encounter at each station, weighted by the coefficient γ, plus the in-vehicle travel time incurred by on-board passengers due to the slack times.

Although difficult to prove we suspect that the general heterogeneous case with multiple degrees of freedom has a unique optimum. This was tested numerically by attempting to numerically solve the optimization problem using multiple starting points within the stability region for the control. In all cases the obtained solution was unique for a given control method. This optimization is solved using the numerical BFGS algorithm [15], where the gradient of the objective function is obtained via finite differences.

If the service operator is interested in guaranteeing a particular reliability level in terms of deviations from schedule, the mathematical program MP4.1 would have to consider additional constraints to reflect the requirements of that goal. The following mathematical program is the result:

$$\text{(MP4.2)} \qquad \min_{\boldsymbol{f}} \frac{\sum_{s=0}^{S-1}[(\gamma b_s(\boldsymbol{f})\frac{H(\boldsymbol{f})}{2}(1+\frac{\sigma_{h,s}^2}{H^2(\boldsymbol{f})}))+(o_s(\boldsymbol{f})d_s(\boldsymbol{f}))]}{H(\boldsymbol{f})\sum_{s=0}^{S}\lambda_s}$$
$$\text{s.t. } d_s(\boldsymbol{f}) = 3\sigma_{D,s}(\boldsymbol{f},\beta_s),$$
$$\sigma_{\varepsilon,s} \leq s_\varepsilon,$$
$$F \leq 1.$$

In this particular single line system with 15 stations the optimization could be solved in a matter of seconds after approximately 50 to 70 iterations of the BFGS algorithm, depending on the type of control method. In more complex scenarios the computation time will increase but this should not be a concern since the optimization is performed off-line.

4.4 Results and Discussion

This section presents the optimization and simulation results obtained in the consolidated Perimeter line. The control coefficients for the different linear holding methods are optimized with MP4.1 using $\gamma = 2$, a value consistent with other studies [76]. Those coefficients are then used as inputs in a battery of 100 simulations of the consolidated Perimeter line. Table 4.1

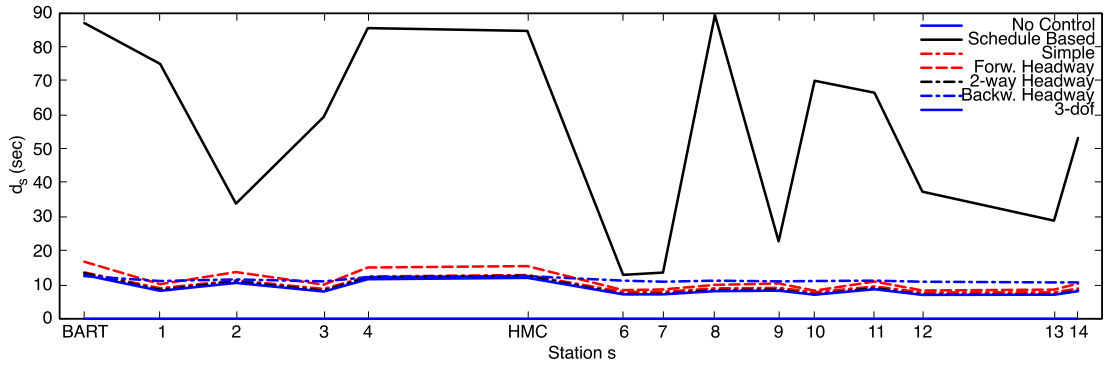
shows the predicted performance, including the value of the objective function, the average slack time $\bar{d}_s$, the average standard deviation of the deviations from schedule across all stops $\bar{\sigma}_\varepsilon$, and the average standard deviation of the headways across all stops $\bar{\sigma}_h$, for all the linear dynamic holding control methods introduced of Section 3.1.2, as well as a control with three degrees of freedom.

Table 4.1: Predicted performance and optimal control coefficient and slack time selection, using Perimeter line data and $\gamma = 2$.

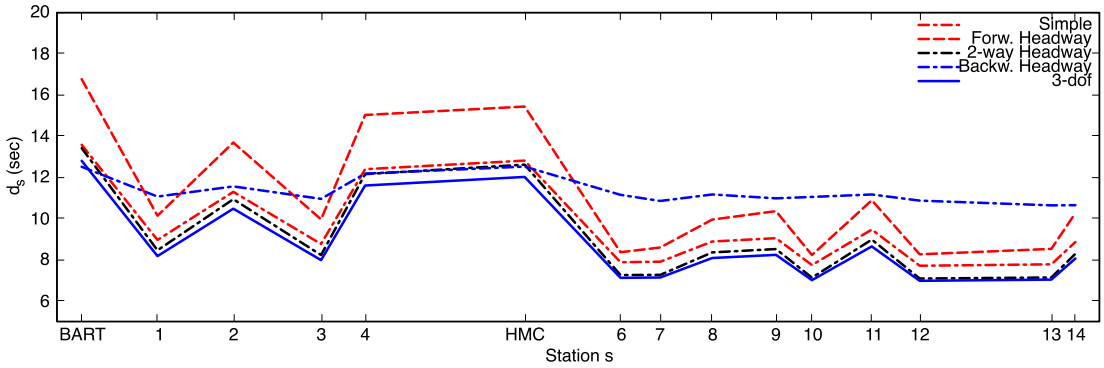
Control Method	O.F. (s)	$\bar{d}_s$ (s)	$\bar{\sigma}_\varepsilon$ (s)	$\bar{\sigma}_h$ (s)	$H(\boldsymbol{f})$	f_{-1}	f_0	f_1
Schedule-based	1,128.0	54.6	18.0	25.5	8:28	0	0	0
Simple Control	485.7	9.5	84.4	119.3	5:34	0	0.97316	0
Forward headway	514.0	10.9	109.1	130.2	5:39	0	0.98448	0.01552
Backward headway	520.5	11.2	104.0	129.1	5:41	0.02858	$0.97142 + \beta_s$	$-\beta_s$
Two-way headway	474.9	9.0	101.5	114.8	5:32	0.011063	0.97787	0.011063
Unconstrained 3-dof	473.4	8.8	99.2	118.5	5:31	0.0120164	0.979049	0.0051439

As expected, the best optimization results are obtained by the control method with three degrees of freedom, where the control coefficients $\{f_{-1}, f_0, f_1\}$ are freely chosen. This minimizes the objective function with a value of 473.4 seconds. The two-way headway-based control results follow closely, with a 0.3% difference in the average travel time that depends on the control choice. The simple control method also achieves a similar performance, differing only 2.6% from the best results presented in Table 4.1. This is consistent with the notion introduced in Section 3.3 that the simple control can achieve quasi-optimal results. Finally, the optimized choice of the forward and backward headway-based strategies would lead to objective function values that are 8.6% and 9.9% higher respectively. All these results are in line with the analysis of a homogeneous bus line. In fact, the low demand area in Figure 3.4 ($\beta \approx 0.01$) presents the exact same order among the different linear dynamic holding control methods that was obtained in this heterogeneous and more realistic framework.

Another interesting aspect of the optimization results are the specific choices of slack times for each of the different methods. Table 4.1 reveals slight differences in the average slack values required by these methods (with the exception of the schedule-based method), but the aggregate nature of the results in the table does not allow to see if the slack values per stop are similar or if they differ greatly between control methods. Figure 4.5 shows the specific slack times at all stops required in each of the simulated scenarios.



(a) All simulated scenarios.



(b) Detail view: Headway-based, simple, and unconstrained 3-dof methods.

Figure 4.5: Optimally callibrated slack times per stop, d_s, for a variety of dynamic holding methods.

Figure 4.5a reveals the disparity existing between the slack requirements that a purely schedule based control would impose if it is applied at all stops versus any of the other strategies. The backward headway-based method does not differ in terms of magnitude from the other methods, but it does in its stop by stop values. The specific form of the control coefficients $\{f_0, f_1\}$ in the backward headway-based method includes the unitary demand parameter β_s at each station. This takes away a certain degree of flexibility in the choice of the control coefficients and results in this profile difference. The rest of strategies have similar control coefficient values with $f_0 \approx 1$ and $f_{-1} \approx f_1 \approx 0$.

After determining the control coefficients and slack times, the next step is to assess the performance of these methods using the event-based simulation. Table 4.2 compares the simulation results obtained by the different control methods. The individual simulations include a warm-up period of 30 minutes, which is approximately the length of a bus cycle in all scenarios except for the schedule-based control. This allows the vehicles and passenger queues to get into conditions that are typical of a steady-state in the system. The following 2 hours of operation are used to collect data.

The results are presented using various performance metrics. These metrics are divided

into two categories. The first category is related to speed, and includes commercial speed (in kilometers/hour) and holding time percentage. Commercial speed is the average speed of the buses, including not only cruising time between stops but also passenger boarding and holding times (if applicable). Holding time percentage is the percentage of time that buses are held at a station for control purposes, out of the total operation time of all buses.

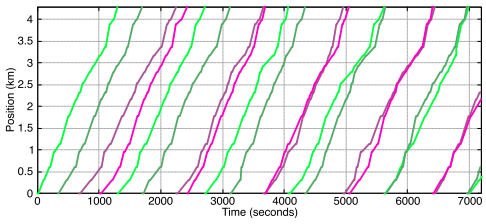
The second category of performance metrics relates to service reliability and it includes the standard deviation of headway (in seconds), the standard deviation of schedule deviations (earliness/delays, in seconds), on-time percentage, bunching percentage, and headway adherence. The standard deviation of headways ($h_{n,s}$) and schedule deviations ($\varepsilon_{n,s}$) are calculated across all buses and all stations. The on-time percentage is defined to be the percentage of bus arrivals that are less than one minute early and less than five minutes late ($\Pr\{\varepsilon_{n,s} > -1 \text{ min } \& \ \varepsilon_{n,s} < 5 \text{ min}\}$), per [12].[3] The bunching percentage is defined to be the percentage of bus headways that are less than one minute apart ($\Pr\{h_{n,s} < 1 \text{ min}\}$), per [8]. Headway adherence is defined by [7] to be the ratio of standard deviation of headway deviation over scheduled headway ($\text{std}(h_{n,s} - H)/H$).

Table 4.2: Simulated performance metrics, optimized linear dynamic holding control methods and uncontrolled scenario.

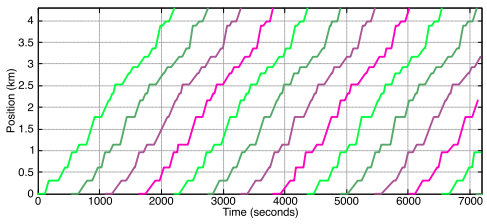
Scenarios	Commercial speed (km/hr)	Holding time %	$\bar{\sigma_\varepsilon}$ (s)	$\bar{\sigma_h}$ (s)	On-time %	Bunching %	Headway adherence
Uncontrolled	11.48	0.00	239.06	256.56	49.73	16.63	0.76
Schedule-based	7.06	20.44	24.72	29.78	99.7	0.00	0.05
Simple Control	10.35	9.60	83.21	117.02	75.00	0.35	0.31
Forward headway	10.21	10.76	105.21	132.21	64.15	0.70	0.35
Backward headway	10.20	11.04	103.47	129.02	62.58	0.36	0.34
Two-way headway	10.40	9.00	100.79	110.28	73.07	0.54	0.30
Unconstrained 3-dof	10.44	8.68	95.87	121.43	83.12	0.38	0.32

Not surprisingly, the uncontrolled scenario has the highest speed among all strategies, because no holding is applied to any buses. However, this particular outcome needs further explanation. First, in real world bus systems, even if no control is applied at intermediate stops, the bus schedules generally include a large slack time or break time at the end of a bus cycle. Transit agencies do this as a simple but somewhat naive approach to compensate for disturbances and allow delayed buses to recover. In those cases, the commercial speeds would certainly be lower. Second, the simulation allows for bus leapfrogging (or overtaking) which is not always possible or allowed. If bus overtaking is limited, the commercial speed

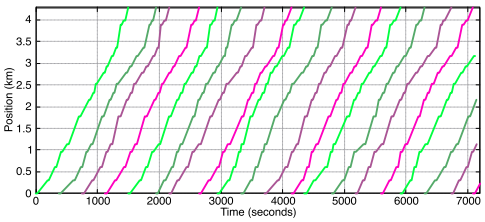
[3]Note that this metric is calculated under the assumption that the virtual schedule used in the control and the published schedule used by the passengers coincide. In reality, an agency could shift the published schedule in a way that would maximize the on-time performance of the system by generally delaying all buses a bit. Under this approach, the strategy with the lowest earliness/lateness standard deviations σ_ε would obtain the best on-time results.



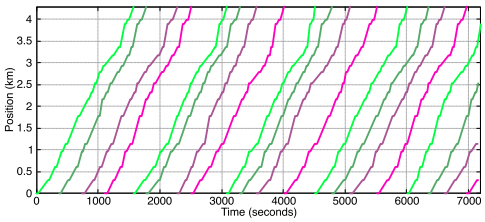
(a) Uncontrolled scenario.



(b) Schedule-based control.



(c) Simple control.



(d) 3-dof control.

Figure 4.6: Simulated bus trajectories obtained from a single simulation run for different control scenarios.

would also decrease because of a larger incidence of bunching, resulting in longer boarding times for the bunched buses. Third, the results in the uncontrolled case are limited by the time scope of the simulation. If the simulation period was 4 hours long instead of 2, the reliability results in the uncontrolled scenario would worsen because the system is unstable, see Figure4.6a.

The best reliability results are achieved by the application of schedule-based control at all stops, see Figure 4.6b. However, the burden that this sort of control imposes on the commercial speed makes this approach completely unattractive both for transit users and operators. Transit users would face excessively long trip times. Transit operators would need to use more vehicles to provide a similar level of service in terms of frequency. These drawbacks make the application of this strategy a no-go.

The remaining control methods provide an excellent compromise between reliability levels and commercial speed (ranging from 10.20 km/hr to 10.44 km/hr), see Figures4.6c and 4.6d. The control with three degrees of freedom yielded the best commercial speed among the controlled scenarios. This is reasonable since the objective function that was minimized to calibrate the control coefficients modeled the average travel time. The values of $\bar{\sigma}_\varepsilon$ and $\bar{\sigma}_h$ for all the control methods are consistent with the numerical calculations provided in Table4.1 and are also fairly close among these methods. The simple control yielded the smallest $\bar{\sigma}_\varepsilon$ among the non schedule-based methods, with 83.2 seconds. The application of the two-way headway control resulted in the lowest $\bar{\sigma}_h$ with 110.28 seconds.

This section showed how the control coefficients can be calibrated to optimize a certain

objective, but the possibilities are endless. The transit agency should adapt the objective function to match their specific goals (e.g. minimizing bus cycle time subject to a reliability constrain or maximizing commercial speed in a specific stretch of the route). Then it should use field data in conjunction with the equations presented in Section 4.3 to determine the specific choice of control coefficients. This particular example was limited to a specific time range in the Perimeter line (10am - 12am, weekdays). In a real world application, the transit agency would need to discretize their operation into different time-frames with similar demand and traffic conditions (e.g. one hour time-frames along the day or morning/evening peaks and off-peak periods). The optimization problem would need to be solved in each one of these time periods yielding a different menu of control coefficients to be applied depending on the time of day.

Chapter 5

Multi-line Corridor Control

The previous chapters have focused on the application of dynamic holding control to isolated bus lines. Unfortunately, considering only isolated lines does not solve all the problems in the vast majority of bus networks in the world. For dynamic holding bus control to be effective in most real scenarios, its scope has to be expanded to corridors with overlapping lines. To the author's knowledge, no article within the bus control literature has yet addressed this problem. Outside the control field, there exist other areas that include scenarios where several bus routes overlap in a corridor, for example in the case of network design problems. In most of these cases, researchers optimistically assume a given level of reliability, despite the lack of effective methods to provide it.

The goal of this chapter is then first to model the motion of the buses in corridors where multiple lines overlap. As we will later see, the presence of multiple lines increases the types of passengers who may be interested to board a bus of a specific line. This chapter then explores the formulation of dynamic linear holding control methods for such scenarios. It is proven that the application of certain control results in buses that behave as if their lines were isolated. Thus, the stability conditions derived for an isolated bus line in Chapter 3 also hold in this more complex system. The application of such control will be tested in a later chapter with field data – first in a simulated environment similar to the one used in Chapter 4, and then in real life with a case study of the system in San Sebastián, Spain. The case study demonstrates that the proposed bus control method is applicable with cheap existing technology.

5.1 Formulation: Two Line Corridor

When multiple lines overlap in a single corridor, modeling the motion of the various types of buses becomes more complex. This type of scenario includes degrees of freedom that do not appear in isolated bus lines. First, when a bus stops at a station to pick up passengers new types of demand have to be considered. In an isolated line, one only needs to know the arrival rate of passengers at that station, since all those passengers can only board one type

of bus: the bus whose line covers that station. With multiple lines one can find passengers who specifically need to board a certain line and passengers who could indifferently board buses from various line combinations. If this aspect is ignored, it is possible to miscalculate the expected bus dwell times.

In addition, the concept of a bus headway also changes when multiple lines overlap. In a corridor with a single line, the concept of headway is unambiguous. The term refers to the time between consecutive arrivals of buses in that particular line. On the other hand, when two lines overlap the term is ambiguous because one can think about two types of headways: intra-line headways (time between two consecutive bus arrivals from the same line at the same station) and inter-line headways (time between two consecutive bus arrivals from different lines at the same station). This differentiation is important when formulating the buses' motion because dwell times then depend on the vector of headway types.

It turns out that the degree of complexity of the multiline corridor problem increases with the number of lines that overlap. In fact, as we will later see, the types of headways and demands grow faster than factorial with respect to the number of lines. In view of this, the two-line scenario is formulated first, and then we extend the formulation to the n-line case.

Virtually all real-world scenarios where two lines overlap once can be schematically represented as in Figure 5.1. In some cases, the individual segments could be non-existent but this figure encompasses all possible two-line scenarios with a single overlapping corridor. The motion of the buses in the individual segments can be characterized using the equations that were originally presented in Section 3.1.2. For this reason, this section focuses exclusively in the overlapping corridor. For simplicity, it is assumed that the central corridor is infinitely long and that buses depart deterministically from the station where the two lines start to overlap. This assumption is realistic since it is possible to control the buses along the initial segments of the two lines where there is no overlap.[1]

The rest of the assumptions that defined the cruising, dwelling and system noise in Section 3.1.1 are almost identical in this section. Only assumption (iii) needs a slight modification to account for the presence of two different lines. Instead of having a unique demand type, it is assumed that the two demand types are known at every station over time. This means that the average line-specific demand rates (arrival rates of passengers who can only board buses from one line) as well as the common demand rate (arrival rate of passengers who can board either one of the two lines) are known.

Note that in reality, if the bus system had initial and final individual line segments, as in Figure 5.1, an additional assumption is necessary. Passengers who board in the initial individual segment of the line could need to transfer to the other line if their destination is in that other line's final individual segment. These passengers would increase the bus dwell times at

[1]To make sure that this assumption holds, the transit agency could operate the simple control method throughout both initial individual segments and then apply the schedule-based control at the station where the lines converge, effectively setting a time point. Therefore all the buses would leave the initial station on schedule without imposing an excessive burden to the passengers that board in the initial segments and travel past the convergence point.

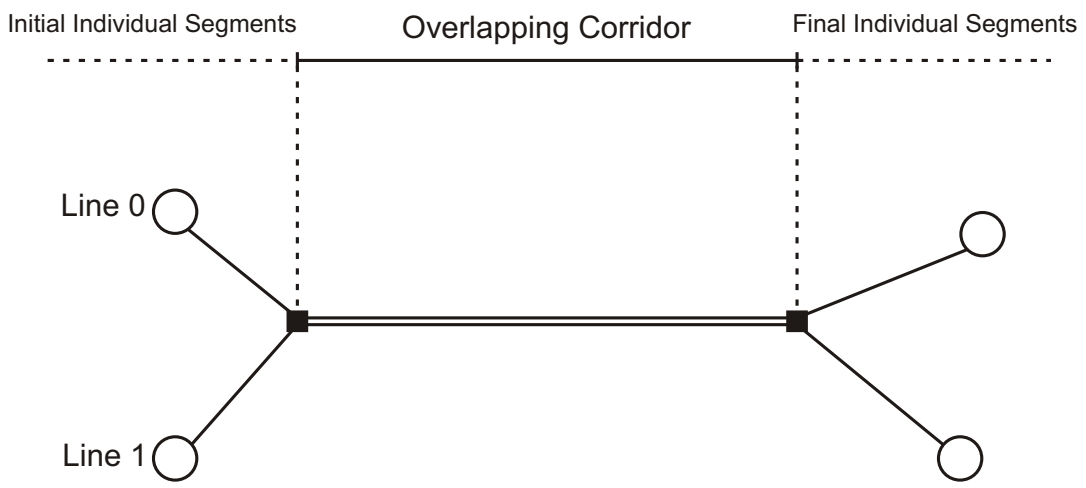


Figure 5.1: Possible two line configurations.

the transfer stops. However, one could assume that those passengers would normally transfer either at the first or the last station of the overlapping corridor, having no effect on the dwell times of intermediate stations. This assumption is reasonable because, by transferring at the first station in the overlapping corridor, standing passengers would generally improve their chances of finding an empty seat on their transfer bus. On the other hand, if a passenger has already found a seat, she would probably prefer to remain seated. Moreover, the transit operator could promote that behavior by encouraging passengers to transfer at those two extreme stations. In the analysis that follows, transfers at the intermediate stations of the overlapping corridor will be assumed not to occur.

Under these assumptions, the bus loading times will only be affected by those passengers who board in the overlapping corridor. As mentioned earlier, there are two kinds of passengers that fit this category: passengers whose destination lies within the overlapping corridor and therefore can board either line, and passengers with destinations in one of the lines final segments who need to use a specific line. To distinguish between them let us introduce the superscripts $l \in \{0, 1\}$, which denotes a specific bus line, and L which denotes the combination of the two lines.[2]

As in the isolated line case, the subscript s will denote the station index. The value $s = 0$ denotes the initial station in the overlapping corridor. The index increases monotonically by 1 for each station in the direction of travel. The notation to index the buses is more complex now. Note that now it is possible to define 3 different sequences of buses at any given station and their index sets. Two that include only the specific buses of one line N_l and another one that includes all buses in the system N_L. The subscript n_l denotes the bus index

[2]In the two line scenario, where $l \in \{0, 1\}$, L can only take one value: $L = \{01\}$.

within the sequence N_l of buses that belong to line l. This index increases monotonically by step 1 and $n_l = 0$ would indicate the first bus of line l to visit a station. On the other hand, the subscript n_L denotes the bus index within the sequence N_L of buses that belong to the combination of lines L. We also assume that there exists a function T_L that maps the index n_l of a bus of line l into the bus' index in the sequence N_L, $T_L(n_l) = n_L$. With these assumptions in mind, a general two-line scenario formulation can be obtained by considering the following notation:

- $t_{n_l,s}$ is the scheduled arrival time of bus n_l at station s,
- $t_{n_L,s}$ is the scheduled arrival time of bus n_L at station s,
- $t_{n_l-1,s}$ is the scheduled arrival time of the bus from line l immediately preceding bus n_l at station s,
- $t_{n_L-1,s}$ is the scheduled arrival time of the bus from either line immediately preceding bus n_L at station s,
- $a_{n_l,s}$ is the actual arrival time of bus n_l at station s,
- $a_{n_L,s}$ is the actual arrival time of bus n_L at station s,
- $a_{n_l-1,s}$ is the actual arrival time of the bus from line l immediately preceding bus n_l at station s,
- $a_{n_L-1,s}$ is the actual arrival time of the bus from either line immediately preceding bus n_L at station s,
- $\nu_{n_l,s+1}$ is the random noise in the trip time of bus n_l between stations s and $s+1$, whose mean is zero and variance is $\sigma^2_{n_l,s+1}$,
- $D^l_{n_l,s}$ is the holding time applied to bus n_l at station s,
- $d^l_{n_l,s}$ is the slack time included in the schedule of bus n_l at station s,
- β^l_s is the unitary demand rate of passengers that need to board line l at station s,[3]
- β^L_s is the unitary demand rate of passengers that can board either line at station s.

In all the formulas below it is assumed that n_l and n_L refer to the same bus; i.e. that $n_L = T_L(n_l)$.

[3]For convenience and conciseness, we assume that demand does not vary over time. Note that this could be easily amended by including a dependency on the index n_l and considering the expected demand rate at a particular time of the day.

The scheduled motion of bus n_l is now formulated using this notation. The scheduled arrival of bus n_l at station $s+1$ is:[4]

$$t_{n_l,s+1} = t_{n_l,s} + \beta_s^l(t_{n_l,s} - t_{n_l-1,s}) + \beta_s^L(t_{n_L,s} - t_{n_L-1,s}) + d_{n_l,s}^l + c_{n_l,s}, \tag{5.1}$$

where $t_{n_l,s} = t_{n_L,s}$. This equation includes an additional term that did not appear in equation (3.1). This term, $\beta_s^L(t_{n_L,s} - t_{n_L-1,s})$, accounts for the expected boarding time of passengers who are indifferent and can board either line. Similarly, the actual arrival time of bus n_l at station $s+1$ can be expressed as:

$$a_{n_l,s+1} = a_{n_l,s} + \beta_s^l(a_{n_l,s} - a_{n_l-1,s}) + \beta_s^L(a_{n_L,s} - a_{n_L-1,s}) + D_{n_l,s}^l + c_{n_l,s} + \nu_{n_l,s+1}, \tag{5.2}$$

where $a_{n_l,s} = a_{n_L,s}$.

Again, as in the isolated line case, we can subtract equation (5.1) from (5.2) and obtain a state equation in terms of the deviations from the schedule. The result is:

$$\varepsilon_{n_l,s+1} = \varepsilon_{n_l,s} + \beta_s^l(\varepsilon_{n_l,s} - \varepsilon_{n_l-1,s}) + \beta_s^L(\varepsilon_{n_L,s} - \varepsilon_{n_L-1,s}) + (D_{n_l,s}^l - d_{n_l,s}^l) + \nu_{n_l,s+1}. \tag{5.3}$$

Recall that $D_{n_l,s}^l$ is a function that depends on the buses' deviations from schedule, such that $D_{n_l,s}^l = d_{n_l,s}^l$ when all these deviations are zero.

Dynamic holding control formulation The formula for $D_{n_l,s}^l$ is now specified for each of the lines $l \in \{0,1\}$ so that equation (5.3) decomposes by line, and remains linear with white noise. The idea is to counteract the effects that demand imposes on the motion of the buses with the holding time. To obtain the desired formula for $D_{n_l,s}^l$ express it generally as a function of the schedule deviations:

$$D_{n_l,s}^l = d_{n_l,s}^l + \sum_i \gamma_i \varepsilon_{n_l-i,s} + \sum_{i'} \rho_{i'} \varepsilon_{n_L-i',s}. \tag{5.4a}$$

Note that the two summation terms have different indexes, to reinforce the idea that they refer to two different bus sequences: buses that belong to line l and buses that belong to either line. This control can be conveniently expressed in the following form:

$$\begin{aligned} D_{n_l,s}^l = d_{n_l,s}^l &- [(1 + \beta_s^l + \beta_s^L)\varepsilon_{n_l,s} - \beta_s^l \varepsilon_{n_l-1,s} - \beta_s^L \varepsilon_{n_L-1,s}] \\ &+ \sum_i f_i^l \varepsilon_{n_l-i,s} + \sum_{i'} f_{i'}^L \varepsilon_{n_L-i',s}, \end{aligned} \tag{5.4b}$$

[4]Note that, contrary to Section 3.1.2, this formulation has dropped the notion of a headway. This is intentional, since this formulation covers a generalized scenario, where inter-line and intra-line headways could differ from one station to another.

where the term $[(1+\beta_s^l+\beta_s^L)\varepsilon_{n_l,s}-\beta_s^l\varepsilon_{n_l-1,s}-\beta_s^L\varepsilon_{n_L-1,s}]$ cancels the demand terms of (5.3). Note that all the information necessary to determine the canceling term is available upon the arrival of bus n_l, since the canceling term only includes deviations that occurred at station s prior or concurrent with the arrival of bus n_l. The vectors $\boldsymbol{f}^l$ and $\boldsymbol{f}^L$ are called "control coefficients". As desired, insertion of (5.4b) into (5.3) results in a simple linear equation with white noise:

$$\varepsilon_{n_l,s+1}=\sum_i f_i^l\varepsilon_{n_l-i,s}+\sum_{i'} f_{i'}^L\varepsilon_{n_L-i',s}+\nu_{n_l,s+1}. \tag{5.4c}$$

And if one chooses $f_i^L=0$ the equation decomposes by line resulting in:

$$\varepsilon_{n_l,s+1}=\sum_i f_i^l\varepsilon_{n_l-i,s}+\nu_{n,s+1}. \tag{5.4d}$$

This equation is of the same form as (3.4c). Therefore all the results of Section 3.1.2 apply to this form of control. In essence, the proposed control transforms the corridor into a superposition of independent lines.

5.2 Formulation: Extension to Any Number of Lines

The complexity of the formulation grows dramatically with the number of lines that overlap in a given corridor. To illustrate this, instead of two lines, imagine that the corridor of interest contains three lines, $l\in\{0,1,2\}$. Then, the total number of possible demand types would be 7: $\{0,1,2,01,02,12,012\}$. This number is obtained by adding the number of 1, 2, and 3-line combinations that can be formed with a set that contains a total of three lines. If we consider a general case with m lines instead, the total number of possible demand types at a station during a certain time interval is 2^m-1.

Regardless of this increased complexity, if the operator has knowledge of the average demand rates, it is still possible to model the motion of the buses and to formulate a canceling holding time that leads to a system behavior as the one obtained in the previous section. To simplify the formulation, let $\mathbb{L}^k$ denote the set of line combinations containing k lines, L be a specific line combination, and β_s^L the demand for that combination at station s. Then the state equation that describes how deviations from schedule propagate through the corridor for bus n_l can be written as:

$$\varepsilon_{n_l,s+1}=\varepsilon_{n_l,s}+\sum_{k=1}^{m}\sum_{L\in\mathbb{L}^k}\beta_s^L(\varepsilon_{n_L,s}-\varepsilon_{n_L-1,s})+(D_{n_l,s}^l-d_{n_l,s}^l)+\nu_{n_l,s+1}. \tag{5.5}$$

As before, the subindex n_l and n_L, $\forall L$, refer to the same bus. Also as with two lines, the holding time can be conveniently formulated to cancel most of the terms in equation (5.5) and greatly simplify the analysis of the noise propagation. Our choice for holding time is:

$$D^l_{n_l,s} = d^l_{n_l,s} - [(1 + \sum_{k=1}^{m} \sum_{L\in\mathbb{L}^k} \beta^L_s)\varepsilon_{n_l,s} - \sum_{k=1}^{m} \sum_{L\in\mathbb{L}^k} \beta^L_s \varepsilon_{n_L-1,s}] + \sum_{k=1}^{m} \sum_{L\in\mathbb{L}^k} \sum_{i} f^L_i \varepsilon_{n_L-i,s}, \tag{5.6a}$$

or even more simply:

$$D^l_{n_l,s} = d^l_{n_l,s} - [(1 + \sum_{k=1}^{m} \sum_{L\in\mathbb{L}^k} \beta^L_s)\varepsilon_{n_l,s} - \sum_{k=1}^{m} \sum_{L\in\mathbb{L}^k} \beta^L_s \varepsilon_{n_L-1,s}] + \sum_{i} f^l_i \varepsilon_{n_l-i,s}. \tag{5.6b}$$

Equation (5.6a) leads to the desired linear equation of motion with white noise and equation (5.6b) reduces the corridor behavior's to a superposition of independent lines. As in the case of two lines, all the information necessary to compute the canceling term $[(1 + \sum_{k=1}^{m} \sum_{L\in\mathbb{L}^k} \beta^L_s)\varepsilon_{n_l,s} - \sum_{k=1}^{m} \sum_{L\in\mathbb{L}^k} \beta^L_s \varepsilon_{n_L-1,s}]$ will be readily available upon the arrival of bus n_l to station s, since only the deviations from schedule from buses that have previously arrived at that station are needed. However, for large m the number of demand parameters $\beta^{\cdot}_s$ that need to be estimated may be impractically large. For this reason an understanding of control methods that do not require so much information should be sought. Regarding the stability of the control method presented suffice to say that the results obtained in Section 3.2 would also apply in this case, since the functional form of the controlled state equations are exactly the same. If the reader is interested in a more detailed analysis of the stability properties and the control calibration, Appendix D includes a deeper look into both aspects.

Chapter 6

Multi-line Case Study: Dbus, San Sebastian

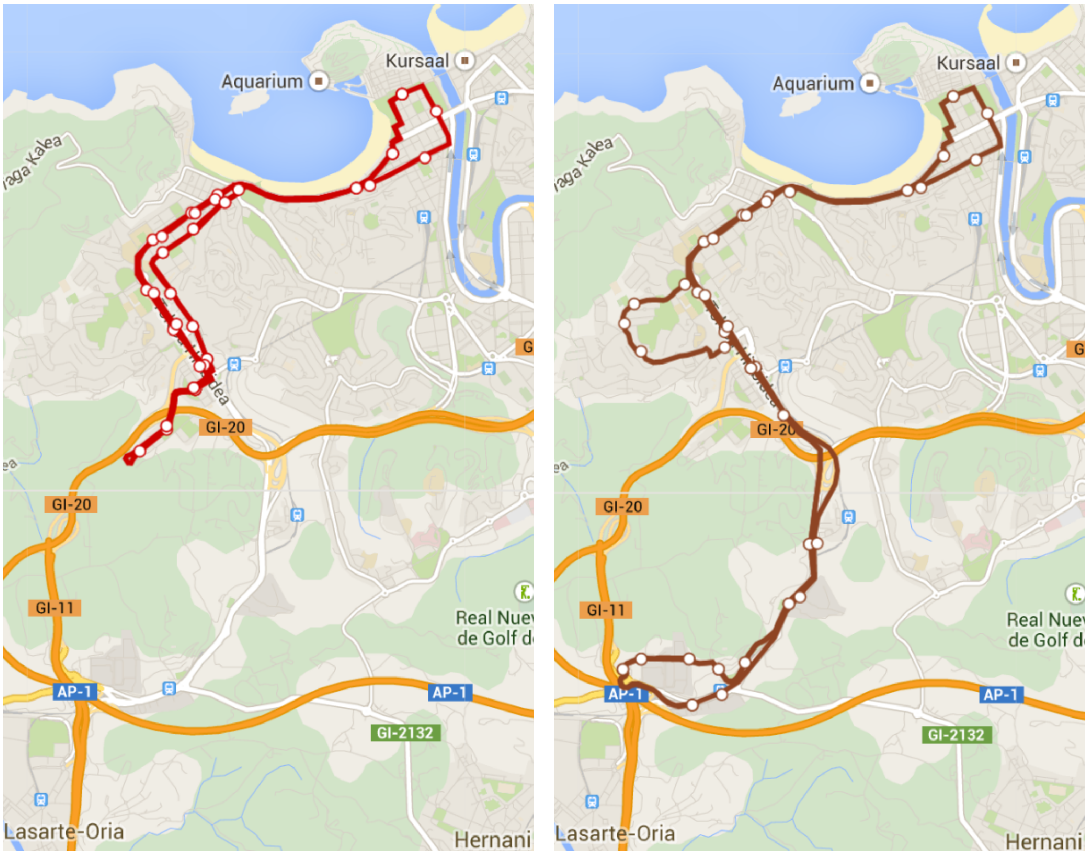
This chapter presents a case study of the multi-line dynamic holding control, which was applied in the transit system of San Sebastián, Spain. The two lines involved in the study, the operational intricacies of the project, the preliminary data collection effort, and the implementation architecture are initially described. The application of the control is first simulated to obtain predicted improvement metrics. These results are then compared with the data collected in the field. The chapter concludes by recounting the practical lessons learned during the implementation of the case study. Particular attention is paid to issues that could affect the control's resilience.

6.1 System Description

Dbus, the transit agency in San Sebastián, Spain, had a complex synchronization problem. Routes 5 and 25, see Figure 6.1, share a long common corridor while also operating independently for long stretches with headways of 6-8 minutes and 20 minutes, respectively.

Dbus wanted to provide the most regular and frequent service possible in the long and busy common corridor taking full advantage of the two lines that serve it. They determined that if all the buses could be precisely kept on a schedule that interleaved the two services they could offer 4-6 minute headways in the corridor. Figure 6.2 illustrates the idea.

Unfortunately, the traffic conditions and high passenger demand for these routes meant that buses frequently deviated from their schedule, delivering irregular service and occasionally pairing. Before the implementation of the control, dispatchers were responsible for the very difficult task of helping drivers maintain the lines synchronized. This was challenging because numerous buses had to be tracked and guided simultaneously using voice communications and because the existing AVL system had low spatio-temporal resolution (vehicles only updated their positions upon arrival to a station). Their available AVL system precluded the detailed observation needed to identify and correct problems before they became



(a) Line 5. (b) Line 25.

Figure 6.1: Map of the two lines controlled in the case study, Dbus, San Sebastián.

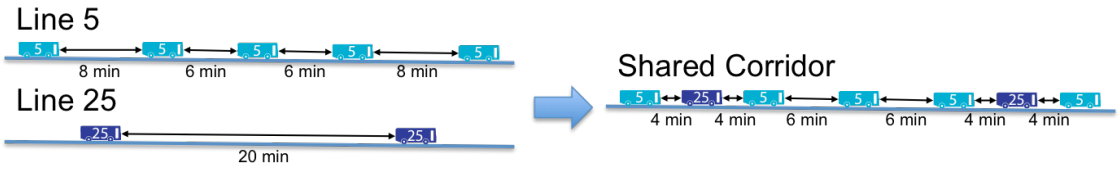


Figure 6.2: Lines 5 and 25, vehicle interleaving diagram.

serious. And with voice commands, dispatchers could not resolve the issues in a timely manner.

With these problems in mind, Dbus decided to apply dynamic holding control with technology that the author and a team of colleagues had developed.[1] This new control approach was intended to provide drivers with the necessary precise and continuous guidance to keep their buses in sync without a dispatcher's assistance.

6.2 Implementation Approach

The system deployed in San Sebastián consisted of a set of on-board devices that had the following essential capabilities: (i) a GPS sensor, (ii) network communications, (iii) a microprocessor, and (iv) a visual display.

The GPS sensor continuously monitors the position of the buses. These measurements are then compared with the vehicle's underlying schedule to obtain the schedule deviation data that is the basic input for the holding time calculations. The GPS sensor in this implementation captures a new data point every second. This high frequency is essential to precisely capture multiple events of interest such as a stop arrival or a departure. Determining when these two events are taking place is essential in order to provide the holding time cues to the drivers at the right time.

The network communications allow the on-board units to relay/retrieve the schedule deviation information to/from a central server. This server receives the schedule deviation data from all the buses in the two lines as soon as it is generated (i.e. upon a vehicle's arrival to a stop) and it makes this information available to all the vehicles in the system using a pub/sub architecture approach [33].

The on-board units dynamically subscribe themselves to different "topics". In this particular application, the topics contain the deviations from schedule collected at specific stops. Thus, a bus that is about to arrive to a stop will subscribe to that stop arrival topic, retrieving the necessary deviations from schedule to calculate its holding time upon arrival, see for example equation (5.4b).

The on-board communications are also used to retrieve the service information for a particular vehicle. In San Sebastián, bus drivers input their trip information every time they start the service in a console located on their dashboard. This information is communicated immediately after to the control on-board units using a direct exchange with the on-board AVL unit. This allows the control on-board units to retrieve the trip's schedule data.[2].

The microprocessor provides the on-board units with enough computing power to carry out diverse functions that are integral to the correct application of the control such as: refining raw GPS measures using the route's known shape, calculating delays, determining the state of the vehicle (e.g. circulating between stops, arriving to a stop, dwelling at a stop,

[1] For more information about the technology go to `http://www.v-a.io`

[2] The schedule data is stored in the on-board device's memory using a database that contains the static schedule information in the GTFS format [40]

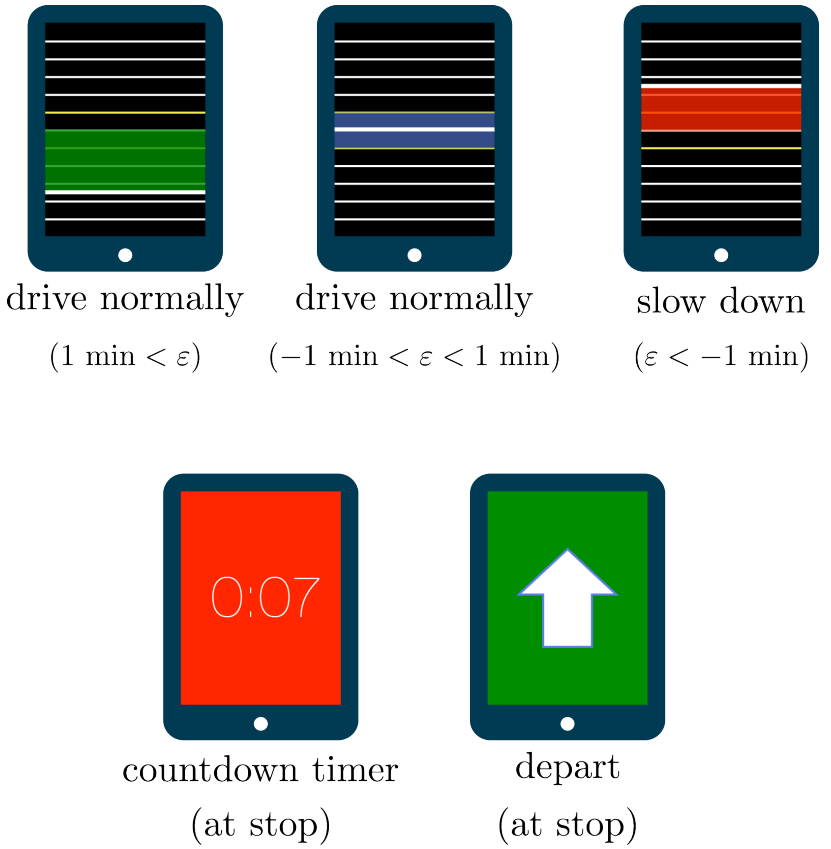


Figure 6.3: User interface implemented in the San Sebastián case study. Cruising guidance (top row) and holding control (bottom row).

departing from a stop, etc.) through a finite state machine approach [48], and computing the holding times.

Finally, a visual display provides feedback to the drivers in real-time. The visual interface deployed has two modes. First, a guidance mode is displayed while the buses are circulating between stops, as depicted in the top row of Figure 6.3. In these circumstances, a vertical bar informs the driver of her current position with respect to the underlying schedule. The height of the bar is directly proportional to the magnitude of the schedule deviation. The bar is color coded based on that deviation. For schedule deviations that exceed one minute of earliness ($\varepsilon < -1$ min) the bar is colored red, to communicate the need to slow down. Schedule deviations that exceed one minute of delay ($\varepsilon > 1$ min) trigger a green bar, indicating the driver that he can proceed normally. Finally, if the absolute value of the deviation from schedule is less than one minute (-1 min $< \varepsilon < 1$ min), the central region of the screen is colored blue.

Second, a holding mode is displayed at the stops using a simple countdown interface, as depicted in the lower half of Figure 6.3. Once the on-board units recognize that the bus has arrived to a stop, the holding time is computed. This time is then shown to the driver

in a red background indicating to the driver that he needs to wait until the countdown is completed. After the countdown ends, a brief animation depicting a vertically sliding white arrow on a green background indicates that the holding time has finished and that the driver can depart. Once the driver accelerates and the on-board unit captures the departure event, the interface automatically switches to the cruising mode. If the driver decides to depart before the countdown is completed the interface also switches automatically to the cruising mode.

Figure 6.4: Installation of the on-board control equipment in Dbus, San Sebastián.

The on-board units finally installed in Dbus were off-the-shelf tablet devices. This equipment satisfies all the above requirements at a very low cost. These tablets were mounted on the buses' frame, within the drivers' field of vision, see Figure 6.4. This installation configuration allows drivers to get feedback with a quick glance without impeding their field of vision.

6.2.1 Parameter Characterization

Before implementing the control in the field, it was necessary to collect data to establish a "before" benchmark, to calibrate the control, and to perform simulations to predict its performance. These data were collected with the on-board devices in a passive state. The screens were turned off but raw GPS position data were collected for the month between March 6th and April 5th, 2013. These data were then processed in conjunction with the underlying schedule to obtain the exact times of arrival and departure from the different stations in the two lines. The data were then used to obtain the noise parameters that affect the motion of the buses between stops and to measure the original reliability levels in the uncontrolled system, which will be presented later.

Noise

In order to obtain the different noise levels, first the service days were divided into the four different types included in Dbus' underlying schedule: Workdays (Monday - Thursday), Fridays, Saturdays, and Sundays/Holidays. For each one of these service days, the time of day was discretized on an hourly basis. Thus, the noise levels collected were not only heterogeneous in space but also in time. All bus travel times for a given hour (e.g. between 8am and 9am) and between the same two consecutive stations were grouped together as a sample. The sample standard deviation for that particular hour h and destination stop s, $\sigma_s(h)$, was then used to estimate the noise level of the group.[3]

Demand

The application of the dynamic holding control in Dbus required the estimation of both line specific and common demand parameters at each one of the stops and for each one of the different hours of operation h ($\beta_s^5(h)$,$\beta_s^{25}(h)$,$\beta_s^{5,25}(h)$)[4]. Unfortunately, the data collected by Dbus' Automated Passanger Counting (APC) system [17] was not retrieved in this format directly.

Dbus had access to the total hourly boarding rates per stop and line, which we will denote as β_s^{5t} and β_s^{25t}. As the reader can imagine, this level of aggregation was not satisfactory for our purposes. By aggregating all passenger boardings to buses in a given line the data loses one degree of resolution. That is, it is not possible to differentiate between those passengers whose only choice was to board a bus from a specific line, such as line 5, or those who were indifferent to any of the two lines. To estimate the split we resorted to a simple route choice algorithm that uses the connectivity and origin/destination matrices of the system. The method is described in Appendix E.

If APC data are not available, the demand parameters can be calibrated by observing the uncontrolled system and regressing the observed dwell times versus the observed headways. More specifically, the dependent variable in the regression model would be the buses' dwell time at the stop. The explanatory variables of the model would be the observed inter- and intra-line headways. The constant coefficient of the regression,would then estimate the lost time that buses have when stopping, i.e. the time it takes for the driver to fully stop and opening/closing the doors. And the coefficients that multiply the various headways would be estimates of the sought β-parameters.

6.2.2 Control Optimization

The control optimization process for this case study had to differ slightly from the approach of Chapter 4. In that case, the optimized parameters were both the control coef-

[3]In bus systems with high frequency and medium demand levels, as it is the case in Dbus, the principal disturbances that buses face is the noise that the buses encounter while cruising between stops.

[4]For convenience and conciseness, in the rest of this subsection we will drop the dependence on h from the notation.

ficients, $\boldsymbol{f}$, and the slack times at the various stops, d_s. This approach assumes that the entity in charge of optimizing the control has the freedom to tweak the buses' schedule appropriately.

In this case study however, Dbus established that the underlying bus schedule could not be modified. Thus, the control calibration was limited to the choice of the control coefficient $\boldsymbol{f}$. This would be problematic if there was not enough slack time in the original schedule, because the system would then be inherently unstable despite the control. Fortunately, Dbus' schedule, as most transit agencies currently do, included enough slack at the end of the lines in order to avoid this issue. But the lack of flexibility in distributing the slack along the line (in the form of optimized $d^l_{n_l,s}$ values) limits the efficacy of the control and restricts the applicability of the theoretical formulation derived earlier to calibrate the control coefficients. To overcome these limitations, we resorted to a two-stage approach where the control was first tested in a simulation environment and then validated in the field.

6.3 System Simulation

The simulation of the two line system is used to select the control coefficients for the Dbus case study. The impossibility of modifying the lines' schedules calls for an alternative avenue in the determination of the control coefficients to be used in the field. The search of control coefficients focuses exclusively on the two-line version of the simple control; see (5.4d). The near-optimal results obtained in previous chapters, as well as the simpler architecture and lower data requirements to implement this type of control, are the key factors that justified this decision.

The simulation framework builds upon the tool described in Chapter 4 but it is modified in two areas. First, the boarding generation now have to consider the multiple types of demand in the system. To do so, every bus arrival to a stop served by both lines is stored in two different arrays: (i) an array containing arrivals from a specific bus line (e.g. arrivals from line 5 only) and (ii) an array that contains all arrivals to that specific stop (arrivals from line 5 and 25). These two arrays were then used to determine the intra- and inter-line headways respectively when a new arrival is simulated. The resulting headways are multiplied by the common and line-specific demand rates at the stop in order to generate the stochastic number of passengers waiting to board. The arrays with the different types of arrivals are also used to determine the holding times since they provide access to the necessary deviations from schedule.

The second area slightly modified inc comparison with the original simulation framework is the schedule determination. In Chapter 4 the schedule was an endogenous parameter. In this case, the controller does not have the possibility of modifying it. For this reason, Dbus' actual schedule was pre-loaded as input data. Each one of the buses that covers routes 5 or 25 incorporates its unique service pattern, using Dbus' GTFS files. The other important aspects of the simulation tool, such as the cruising time generation, follow the approach used previously.

Table 6.1: System-wide average performance metrics for the uncontrolled scenario and the simple control scenario with $f_0 = 0.895$

Reliability Metric	Uncontrolled	$f_0 = 0.895$
$\bar{\sigma}_h$	142.1	123.8
$\bar{\sigma}_\varepsilon$	141.3	121.9
on-time %	73.11	89.24
bunching %	7.52	3.57

The simulation tool was calibrated using field data. The demand values aggregated by line were obtained using Dbus' APC system. These were then refined into three types as explained in Section 6.2.1. The average and standard deviations values for the travel times between consecutive stations resulted from a statistical analysis performed over the one month data sample in which the on-board devices ran passively. The on-board devices monitored the buses arrivals yielding the necessary information to estimate the parameters.

The choice of the control coefficient $\boldsymbol{f} = [\cdots, 0, f_0, 0, \cdots]$ was based on a parameter search over the stable region for the headway deviations, i.e. $f_0 \in [0, 1]$. The operation of the two lines was simulated 100 times for a set of potential values of f_0. After the search, the value of $f_0 = 0.895$ was selected. Table 6.1 compares the controlled and uncontrolled scenarios, both simulated, with average metrics calculated across all stops, hours, and both bus lines.[5]

6.4 Field Results

This section presents before and after results of field study of the inbound common corridor of lines 5 and 25. The before results were obtained between March 6th and April 5th, 2013, when the buses operated uncontrolled. The after period comprises the time between June 22nd to July 21st, 2013, where the simple dynamic holding control was active. The control was first activated in April 6th, 2013, using the coefficient obtained in the previous section. The months of June/July were selected as the after period because of the driver training process. Between April and June, Dbus inspectors were able to train over 90% of their driver population, which includes approximately 400 different employees.

Given that the underlying schedule could not be affected in this implementation, the results focus on metrics that are directly related to the system's reliability and do not consider commercial speed. The metrics of interest are: (i) the on-time performance, (ii) the standard deviation of the deviations from schedule, and (iii) the standard deviation of the headway deviations. All these metrics were defined and explained in Section 4.4. The

[5]Note that this table focuses exclusively on reliability parameters because of the impossibility to modify the underlying schedule, which impedes calibrating the slack times and optimizing a combination of reliability and commercial speed.

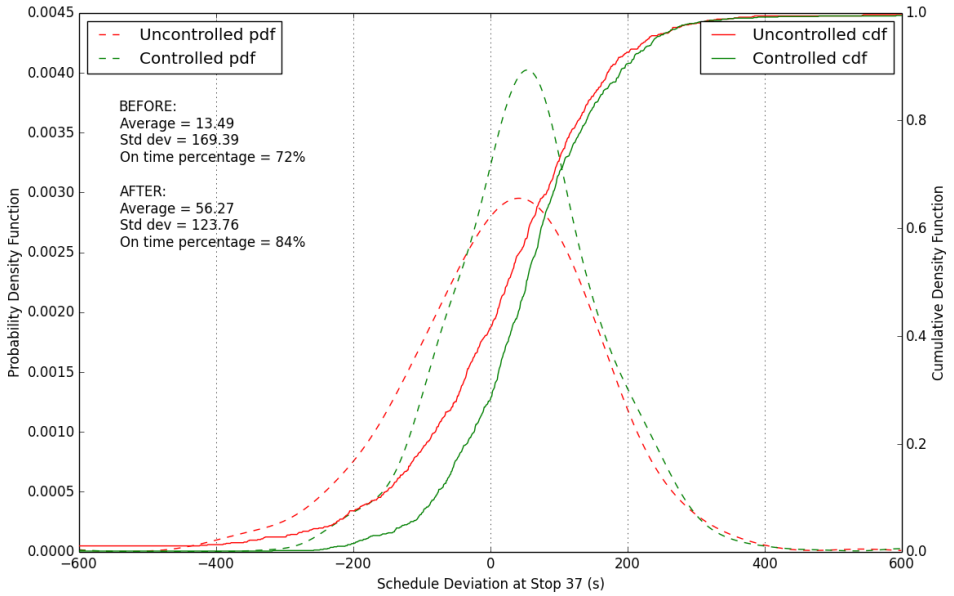


Figure 6.5: Empirical cumulative density function and probability density function for the deviations from schedule at "Okendo 5", Line 5.

results were collected automatically, recording the arrival times and comparing them with the underlying schedule.

Figure 6.5 shows the empirical distributions of the deviations from schedule of the buses in line 5 upon arrival at the "Okendo 5" stop, for the two periods of interest. This is the final stop in the inbound direction and it originally presented the worst reliability results. The reason behind it is that it is the station located the furthest away from the terminal stop, where Dbus has inserted enough slack for the buses to resync. Moreover, this particular direction has the highest demand rates throughout the day. The distributions depicted in the figure were calculated using 845 different data points for the before period and 1,003 for the after period.

Figure 6.5 reveals a significant improvement in the variability of the deviations from schedule. In fact, the application of the simple dynamic holding control was able to reduce the standard deviation of the deviations from schedule from approximately 170 seconds to 124 seconds at that particular stop for bus arrivals in line 5. The on-time performance at the stop also improved from 72% to 84%.[6].

[6]The on-time performance was calculated considering that the virtual schedule used in the the control coincided with the schedule published to the system riders. If the agency wanted to take full advantage of the precise arrival information collected, they could shift the published schedule backwards to increase the on-time performance. In this particular case, by doing so Dbus could have increased the on-time performance at "Okendo 5" to a level as high as 88%.

The shape of the distributions also reveals that one of the main effects of the control in this particular application framework, where the slack could not be added to the schedule, is counteracting the incidence of early arrivals. In the before period the early arrivals at this station accounted for over 41% of the total. In the after period, this value was reduced to approximately 28%. This is a relevant result because in transit systems where the schedule is advertised, early bus departures make users who arrive on-time experience longer wait times. This contradiction between the publicized service and the actual one results in very negative user experiences [18]. Figure 6.6 presents, with a lower level of detail, the superimposed probability density functions of the before and after measurements across all stops and for line 5 and 25 in the inbound common corridor. These distributions were obtained using Gaussian kernel density estimation [67].

The headway deviation comparison also reveal some improvement with control. Figure 6.7 presents a similar detailed view of the results obtained at "Okendo 5" for the common headways, $h^{5,25}$, i.e. the headways between consecutive buses belonging to any of the two lines involved in the case study. The figure shows that the histogram in the controlled case is narrower with smaller tails. This is quantified by the reduction in the standard deviation, which decreases from an original value of 196 seconds in the uncontrolled case to 167.4 seconds.

The line specific results obtained for the inbound overlapping corridor are summarized in Table 6.2. This table includes the on-time performance metrics, the standard deviations of the deviations from schedule σ_ε, and the standard deviations of the intra-line headway deviations σ_{h^5} and $\sigma_{h^{25}}$. The inbound common corridor starts at "America Plaza" and finishes at "Okendo 5", covering ten different stops.

These results show an average reduction of σ_ε of approximately 32% across all the stations in the common corridor. The reduction for buses of line 5 only is 42% while it is 28% for line 25. The absolute increase in the average on-time performance is less dramatic, improving by 12% on average in line 5 and by 5% for line 25. The average on-time performance index in line 5 increased from 82% all the way to 92% across all stops and from 80% to 84% in line 25. However, both results are much higher than usual reported values in North American agencies, which generally perform at the 50-70% level, [55, 4].[7] Finally, the standard deviations of the headway deviations were reduced by 25% and 12% for lines 5 and 25 respectively.

Across all metrics, the improvements in line 25 are smaller because in that case, the common corridor is preceded by a long segment in which the schedule is designed with almost no slack, degrading the reliability performance and reducing the effectiveness of the control. If the schedule could have been modified, distributing some of the slack at the terminal station along the rest of the stops on the line, the results could have been even better.

[7]However, these results could be further improved if the virtual and published schedule were appropriately shifted. In this particular case, Dbus could publicize a schedule that was advanced with respect to the virtual schedule followed by the buses. The value of such shift could be calculated in order to maximize the final on-time performance metric.

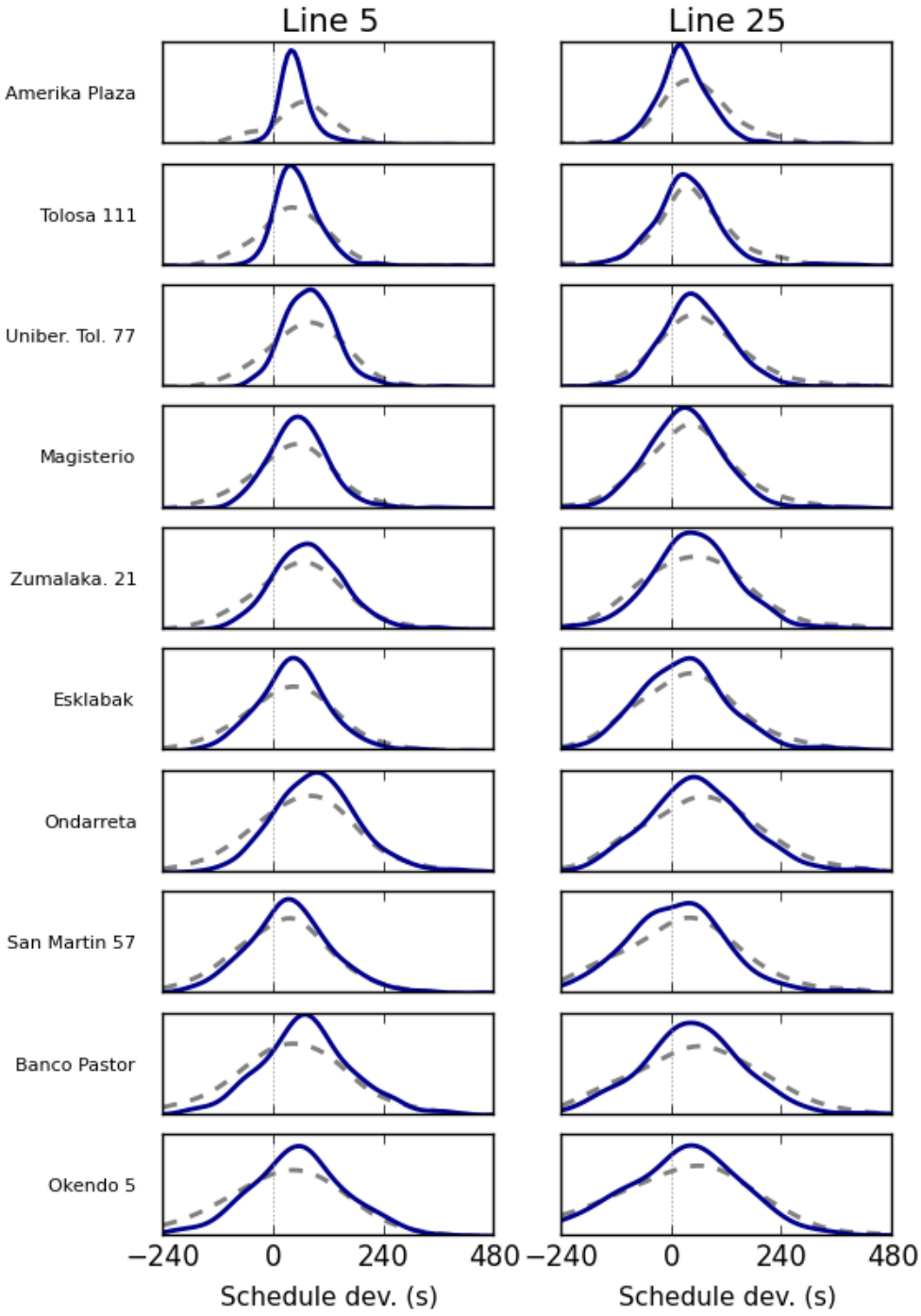


Figure 6.6: Empirical probability density functions of the deviations from schedule along the common corridor for lines 5 and 25. The uncontrolled results are depicted with a dashed gray line while the controlled results are shown with a continuous dark blue line.

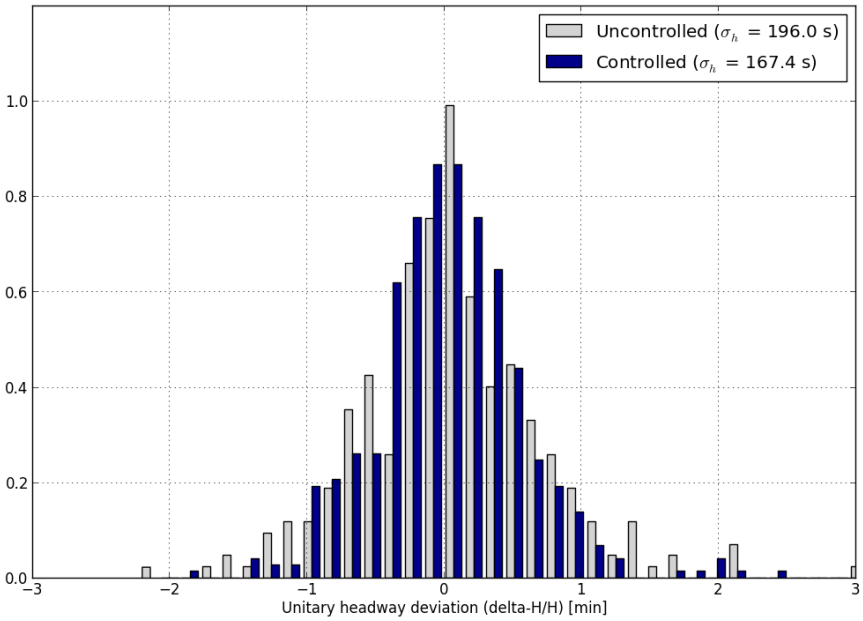


Figure 6.7: Empirical histograms of the unitary headway deviations obtained at "Okendo 5".

Table 6.3 presents the headway reliability results for consecutive buses irrespective of their actual line; i.e. the aforementioned inter-line headways from Chapter 5. The table results show that on average, the standard deviation of inter-line headway deviations was reduced by almost 23% from an original average of 2 minutes and 42 seconds to 2 minutes and 5 seconds. Since lines 5 and 25 provide high frequency service in the corridor, the expected wait times $\bar{w}$ for the users can be computed as $\bar{w} = \frac{\bar{h}^{5,25}}{2}(1 + \frac{\sigma^2_{h^{5,25}}}{(\bar{h}^{5,25})^2})$; see [22]. The expected wait time has two parts: (i) one depending on the underlying service frequency $\bar{h}^{5,25}/2$ (which the control can't affect), and (ii) a second one depending on the system's reliability $\sigma^2_{h^{5,25}}/(2\bar{h}^{5,25})$ (which should be improved by the control). In the inbound overlapping corridor, since $\bar{h}^{5,25} = 6.20$ minutes, the avoidable wait time (ii) was in fact reduced on average by 40%.

The table also includes the percentage of inter-line bunching. Bunching is defined as the occurrence of two consecutive bus arrivals that are less than 1 minute apart. Reducing bunching in these cases was the primary goal for Dbus. The results speak for themselves. The reduction in inter-line bunching across all the stops was 46%.

The result observed in the field did not exactly correspond to the ones obtained via simulation. Table 6.4 presents the equivalent results from Table 6.1 as measured in the field. The bunching and on-time improvements were fairly similar to the ones obtained by the simulation tool. On the other hand both headway and schedule deviations results present

Table 6.2: Schedule adherence results from line 5 and 25 in the stops located along the inbound common corridor.

		BEFORE			AFTER		
Route	[Id] Stop Name	σ_ε (s)	On time %	σ_h (s)	σ_ε (s)	On time %	σ_h (s)
5	[46] Amerika Plaza	125	91	129	53	99	64
	[45] Tolosa 111	125	87	122	55	99	70
	[44] Unibertsitatea Tol.77	133	90	146	68	99	91
	[43] Magisterio	152	85	156	83	96	101
	[42] Zumalaka.21	156	85	155	86	95	119
	[41] Esklabak	166	81	157	103	89	130
	[40] Ondarreta	176	82	178	106	93	135
	[39] San Martin 57	160	74	175	110	83	139
	[38] B.Pastor	170	76	159	103	88	153
	[37] Okendo 5	169	72	177	124	84	166
25	[46] Amerika Plaza	108	91	117	66	93	95
	[45] Tolosa 111	93	89	116	79	90	105
	[44] Unibertsitatea Tol.77	114	89	140	80	94	106
	[43] Magisterio	130	83	131	87	85	125
	[42] Zumalaka.21	131	80	144	100	87	137
	[41] Esklabak	145	77	156	105	78	138
	[40] Ondarreta	164	76	171	114	84	157
	[39] San Martin 57	162	70	168	117	73	159
	[38] B.Pastor	160	75	225	116	81	180
	[37] Okendo 5	195	67	215	149	72	185

Table 6.3: Common headway deviations standard deviations, $\sigma_{h^{5,25}}$, and bunching percentage across consecutive bus arrivals, irrespective of their lines, in the inbound common corridor.

	BEFORE		AFTER	
[Id] Stop Name	$\sigma_{h^{5,25}}$ (s)	Bunching %	$\sigma_{h^{5,25}}$ (s)	Bunching %
[46] Amerika Plaza	138	4.28%	79	0.58%
[45] Tolosa 111	122	2.83%	83	0.79%
[44] Unibertsitatea Tol.77	151	5.63%	104	2.26%
[43] Magisterio	154	4.77%	108	1.79%
[42] Zumalaka.21	165	5.49%	130	3.37%
[41] Esklabak	157	5.36%	133	3.25%
[40] Ondarreta	183	8.33%	148	6.76%
[39] San Martin 57	187	9.11%	149	5.04%
[38] B.Pastor	173	8.57%	154	5.78%
[37] Okendo 5	196	10.45%	167	5.49%

Table 6.4: System-wide average performance metrics measured in the field for the uncontrolled and controlled scenarios.

Reliability Metric	Uncontrolled	Controlled
$\bar{\sigma}_h^{5,25}$	156.9	127.2
σ_ε	146.76	95.15
on-time %	81.06	88.04
bunching %	6.84	3.51

more significant improvements. Multiple reasons could explain these differences. However, we believe that the main contributor is the inclusion of a cruising guidance system in the field. This additional assistance tool for the drivers is not modeled in the simulation. The guidance while driving could have had a significant impact in the reliability improvements captured in the field.

Finally, our field experience revealed that other elements of the system, such as the connectivity of a device or its GPS sensor, could suffer sporadic temporary malfunctions. These factors did not have a significant impact in the overall system performance but it is clearly necessary to account and analyze them. In view of that, the following section will enumerate potential real world aspects that could negatively affect the performance of the control and it will deeply delve into one of them: GPS malfunctions.

6.5 Control Resilience

The implementation of dynamic holding control in San Sebastián proved to be challenging. The theoretical framework developed in earlier chapters includes various simplifying assumptions. These assumptions result in tractable models that yield very relevant insights into the problem. However, in some cases, they do not perfectly reflect the real conditions in the field. This section identifies various issues that arose during the implementation, which were not directly addressed in the theoretical analysis.

6.5.1 Potential Issues

Differences between the theoretical assumptions and the field reality could threaten the practical resilience of the control. For this reason, it is important to identify these differences and analyze their effects with the tools at our disposal. Among the different practical aspects that may pose a risk on final performance, one can find GPS malfunctions, communication issues, power failures, and driver non-compliance. They are now briefly described one at a time.

GPS Malfunctions

The control methods proposed in this dissertation use the buses deviations from schedule upon arrival to a station as their main input. Without this information, the control cannot be reliably applied. Thus the models assume that the controller can continuously monitor the position of the buses, and automatically detect their arrivals to the stations along the line.

The public transportation literature contains numerous articles dedicated to determine how to best use GPS data to track transit vehicles and characterize their operations; see [6, 14, 71] to name a few. In our case, these techniques were adapted to detect the arrivals at the stops using the on-board equipment's GPS capabilities, to compute the deviation from the schedule and to immediately communicate that information.

However, in some occasions the GPS malfunctioned and was not active for a brief period of time. These outages may be limited to a specific geographic area where GPS coverage is consistently bad for all buses, or they may arise unexpectedly and randomly one bus at a time. During outages of either type, it is not possible to produce arrival information for the affected bus. Thus, in the case of the simple control method in an isolated line, two buses in the system will not have the necessary information to compute their appropriate holding times: the affected bus and its follower. This type of issue could be potentially very pernicious.

Fortunately, GPS failures rarely lasted more than 5 to 10 minutes, and there are ways in which the equipment can be reset automatically to function appropriately. However, due to their potential gravity, this chapter will analyze the effects of both GPS outages, and how to treat them appropriately. The bright side is that with the advance of technology, GPS malfunctions should progressively become less frequent and pronounced.

Communication Malfunctions

Another aspect that the control takes for granted is the distribution of the necessary information in a timely manner. In order to apply the control, the buses' on-board devices need access to not just their own deviations from schedule, but also to the deviations from other buses. The system description in Section 6.1 shows how this can be achieved with the right architecture. However, the architecture described relies on the existence of a dependable communication channel.

While implementing the control, there were instances in which the communications of a particular bus did not work. Like the GPS malfunctions, these issues could be associated with a particular area or could arise sporadically. Some regions in the city may have bad or even non-existent wireless internet coverage. In addition, in cases where internet access is provided using on-board WIFI (as was the case with Dbus), the on-board device in charge of providing the service may fail.

The main difference with the GPS issue is that in this case, the affected bus can still generate its own arrival information, although it may not know the most recent information

for other buses. This is important for two reasons. First, the on-board units can store this information locally and make it available as soon as the communications are restored. Second, the affected bus' deviation from schedule is the most relevant piece of information in the application of the dynamic holding control for the affected bus. To see this, simply look at the optimal control coefficients in Section 4.4. The value of f_0 is invariably about 100 times larger than any of the other coefficients. Since this particular coefficient is the one that multiplies the deviations from schedule of the affected bus, it is reasonable to expect communication issues not to be as pernicious as a GPS outage.

Again, as it is the case with the GPS malfunctions, the impacts of communication availability issues should decrease in the coming years. The past decades have seen a continuous improvement of mobile technologies and it seems hard to believe that this trend will be halted in the near future.

Power and Installation Malfunctions

These types of issues arose mainly during the initial stages of the project. Althoug mounting the devices on the buses is a simple task if the right technology is used, small glitches materialized during the first days of the operation. These initial problems were relatively easy to overcome because during the first stages of the operation, the tablets were passively collecting data, without providing any feedback to the drivers. This facilitated changes to the mounts and small maintenance tasks, which were the main concerns at that point.

During the rest of the project, the only problem of this sort was sporadic intermittence with the device's connection to the power supply. Due to the buses' impacts and vibrations, the charging connection would sometimes partially detach. On-board units would then lose the power supply. If no actions had been taken, the devices would have ended up turning off. This problem could potentially have similar effects to GPS outages, since in both cases no reliable control can be applied to the affected bus. Unfortunately, while devices can recover from a GPS outage automatically, a charging issue generally requires human intervention. On the other hand, batteries drain gradually so there is time to react if a power disconnection is detected. For this reason, it is essential to have real-time monitoring capabilities that can alert the operator in case that there is a problem with the charging state. It is key to identify and fix this sort of occurrence as soon as possible.

Driver Compliance and Human Factors

Finally, the theory derived in the earlier chapters assumes that the holding recommendations generated by the control will be precisely followed by the bus drivers. In reality, drivers may not be able to follow the control recommendations with extreme precision. On board passengers, traffic, or other external factors may affect the capabilities of the driver population to exactly follow the recommendations. This type of soft failure could be modeled by slightly increasing the variance of the error terms in the motion equations, e.g. (3.3, 5.3).

More serious is the possibility that specific drivers intentionally decide to not follow the recommendations to hold. Fortunately, the prescriptive nature of the control method, which shows drivers a countdown timer, makes it simple to identify and report this behavior. Driver-specific compliance information should be used to better train the drivers and to work with them to improve their acceptance of the system.

Even though this escapes the scope of this dissertation, future work should assess the different control system features that may have an effect on the drivers decision making and how they perceive the control. This work could increase the overall effectiveness of the control in real life.

6.5.2 Overcoming GPS Malfunctions

As mentioned earlier, GPS malfunctions can be grouped into two categories. The first type includes those that arise because of the lack of GPS coverage in a specific geographic area. These problems affect all buses equally and result in the impossibility to generate arrival information for a segment of the route. The second type (random) is less common but they can be more disruptive because of their unforeseeable nature. These are GPS outages that unexpectedly affect a bus due to problems generally related to the hardware and which cannot be anticipated.

However, one should note that given the distributed nature of the control, a sporadic malfunction in one of the buses should remain fairly localized to the bus in question, and perhaps its immediate neighbors, so its system-wide effect should be quite diluted. Obviously, the affected bus will fare worst but it is to be expected that the rest of the buses in the system should proceed with fair normality.

This section analyzes both types of failures and explores this last hypothesis.

Systematic Area-based GPS Failures

This type of GPS malfunction will affect all buses that traverse a region without GPS coverage. Given the impossibility to generate arrival data, if the controller depends exclusively on the on-board devices' GPS, holding cannot be applied at the stops in the affected region. Fortunately, these types of areas can be easily identified with a brief period of passive GPS data collection.

Once the affected regions and their stops are identified, a possible solution consists in characterizing the sequence of stops without GPS coverage as a unique equivalent stop. This virtual stop should have an equivalent travel time, demand rate, and noise such that it reflects the uncontrolled motion along the stops and links affected by the lack of GPS coverage.

Assume that in a line with S stops, the sequence of stops between $s = s_0$ and $s = s_1$ $(0 < s_0 < s_1 < S - 1)$ have no GPS coverage. Given the lack of arrival information at those stops, the motion of the buses between stops $s_0 + 1$ and $s_1 - 1$ would be uncontrolled. For illustration purposes, also assume that demand and noise is location dependent but time

independent in this line. To obtain the equivalent demand rate and noise one just needs to propagate backwards the deviations from schedule from stop s_1 until reaching stop s_0, under the assumption that the motion will be uncontrolled. To do so, let us use the term $\boldsymbol{u}_s = \boldsymbol{u}_{s|1} = [\cdots, 0, 1+\beta_s, -\beta_s, 0, \cdots]$ to represent the uncontrolled motion vector at stop s; see equation (3.5b) for reference. Thus, we can express the uncontrolled motion between consecutive stations as $\boldsymbol{\varepsilon}_{s+1} = \boldsymbol{u}_s * \boldsymbol{\varepsilon}_s + \boldsymbol{\nu}_{s+1}$, and $\boldsymbol{u}_{s|j} = \boldsymbol{u}_s * \boldsymbol{u}_{s-1} * \cdots * \boldsymbol{u}_{s-j+1}$, $j > 1$. Then the motion of the buses in the affected segment between stations s_0 and s_1 can be expressed as a function of the equivalent uncontrolled motion vector $\boldsymbol{u}^{eq}$ and noise $\boldsymbol{nu}^{eq}$:

$$\begin{aligned}
\boldsymbol{\varepsilon}_{s_1} &= \boldsymbol{u}_{s_1-1} * \boldsymbol{\varepsilon}_{s_1-1} + \boldsymbol{\nu}_{s_1} \\
&= \boldsymbol{u}_{s_1-1} * (\boldsymbol{u}_{s_1-2} * \boldsymbol{\varepsilon}_{s_1-2} + \boldsymbol{\nu}_{s_1-1}) + \boldsymbol{\nu}_{s_1} \\
&= \cdots \\
&= \boldsymbol{u}_{s_1-1|s_1-s_0} * \boldsymbol{\varepsilon}_{s_0} + \sum_{j=1}^{s_1-s_0} u_{s_1-1|j} * \boldsymbol{\nu}_{s_1-j+1} \\
&\approx \boldsymbol{u}^{eq} * \boldsymbol{\varepsilon}_{s_0} + \boldsymbol{\nu}^{eq}.
\end{aligned} \tag{6.1a}$$

Note that, similarly as in Appendix B, if we ignore the second order terms (since $\beta_s << 1$, $\forall s$) then $\boldsymbol{u}^{eq} = \boldsymbol{u}_{s_1|s_1-s_0} \approx [\cdots, 0, 1 + \sum_{s=s_0}^{s_1-1} \beta_s, -\sum_{s=s_0}^{s_1-1} \beta_s, 0, \cdots]$. Thus the equivalent demand rate would simply be the sum of demand rates along the corridor without GPS reception.

$$\beta_{s_1}^{eq} = \sum_{s=s_0}^{s_1-1} \beta_s. \tag{6.1b}$$

The equivalent noise can be obtained similarly. The reader can verify that the resulting noise would still have zero mean. Its variance on the other hand would be given by this expression:

$$\begin{aligned}
(\sigma^{eq})^2 &= \mathrm{var}(\sum_{j=1}^{s_1-s_0} u_{s_1-1|j} * \boldsymbol{\nu}_{s_1-j+1}) \\
&= \sum_{j=1}^{s_1-s_0} \sigma_{s_1-j+1}^2 (\sum_k u_{k,s_1-1|j}^2)
\end{aligned} \tag{6.1c}$$

Using these equivalent demand and noise parameters is the simplest approach to account for areas with lack of GPS coverage. Another option that would overcome this problem consists on equipping the affected stations with devices that could enable geofencing [56], such as a Bluetooth beacon. The on-board devices would then be able to detect the arrival

to a specific stop without the use of GPS data. However, this approach requires the effort to install and maintain additional equipment on the field, which may not be a suitable solution for some transit agencies.

Unforeseen GPS Issues

Experience has shown that the GPS from an on-board unit may fail sporadically for a brief period of time. Fortunately, a malfunction of this kind does not affect the on-board communications, which means that the problem can be communicated to the rest of vehicles in the system and acted upon. Under this scenario, we want to explore the possibility of using all the available information to try to estimate the deviations from schedule of the affected bus while the GPS is not working. These estimated values could then be used instead of the real ones in the formulas for holding times.

To illustrate how the failure could be handled, imagine first that the affected driver is alerted as soon as the GPS malfunction arises. Assume also that, after arriving and boarding passengers at the following stations, the driver presses a button on the on-board device's screen to indicate the start of a holding period and query the device for a countdown. The holding time would then be calculated based on the estimated deviation from schedule, and would then be displayed, emulating the normal function of a controlled bus. However, it is still left to be determined if such an approach would be effective.

The following assumptions are made in order to assess the effectiveness of such approach:

- buses operate in an infinitely long and isolated line with homogeneous demand and noise, β and σ,[8]
- all buses in the system are controlled using the simple control, where $f_i = 0$ for all $i \neq 0$,
- bus n suffers a GPS malfunction between stations s and $s+1$,
- at station s, the deviations from schedule of the buses in the line have reached a steady state, i.e. $\mathbb{E}[\varepsilon_{n,s}] = 0$ and $\text{var}(\varepsilon_{n,s}) = \sigma_\varepsilon$, $\forall n$, as per equation (3.11b),
- the on-board device can identify the moment in which the issue arises and it can immediately transition to the new control approach that uses estimated deviations from schedule,
- deviations from schedule from other vehicles can be precisely measured and communicated to the affected device,
- the affected bus will display holding recommendations based on its expected deviations from schedule at each of the following stations $\{s+1, s+2, \cdots\}$, which will be denoted

[8]Recall that having homogeneous demand along a route can be seen as a worst case scenario, see Section 3.3.1

as $\{\hat{\varepsilon}_{n,s+1}, \hat{\varepsilon}_{n,s+2}, \cdots\}$. These estimations will consider the latest available deviation from schedule of the affected bus, $\varepsilon_{n,s}$, and the known deviations from schedule that would impact the motion of the bus (in the case of the simple control, that is simply the deviations of the bus ahead, $\varepsilon_{n-1,\cdot}$, as illustrated in Figure 6.8).

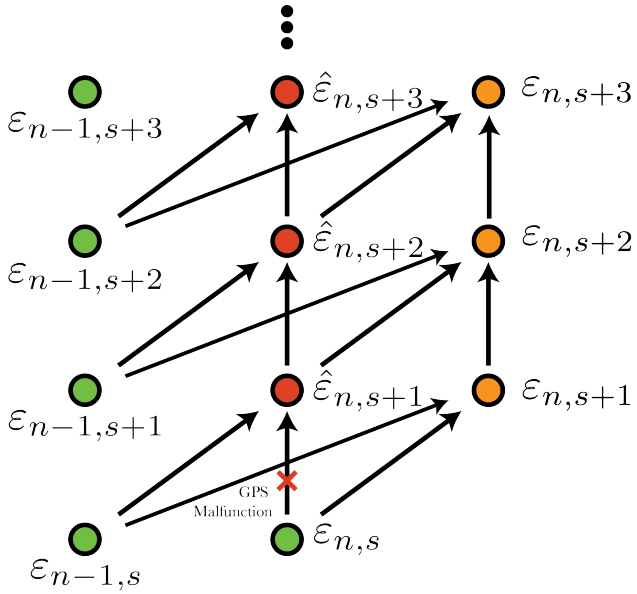


Figure 6.8: Diagram depicting the information flow between the available $\boldsymbol{\varepsilon}_{n-1}$, the estimated $\hat{\boldsymbol{\varepsilon}}_n$, and the not available $\boldsymbol{\varepsilon}_n$ after bus n suffers a GPS malfunction between stations s and $s+1$.

Now let us analyze how using the estimated deviations from schedule instead of the real values would impact the reliability of the affected bus. To do so, we need to iteratively characterize the expected value and variance of the actual deviations from schedule of bus n, along the stations where GPS is not available $\{s+1, s+2, \cdots\}$. At the first affected station $s+1$, the expected deviation from schedule of bus n given the observed deviation from schedule at the previous station is:

$$\begin{aligned}\hat{\varepsilon}_{n,s+1} &= \mathbb{E}[\varepsilon_{n,s+1} \mid \varepsilon_{n,s}] \\ &= \mathbb{E}[\varepsilon_{n,s} + \beta(\varepsilon_{n,s} - \varepsilon_{n,s}) + D_{n,s} - d_s + \nu_{n,s+1}] \\ &= \mathbb{E}[f_0 \varepsilon_{n,s} + \nu_{n,s+1}] \\ &= f_0 \varepsilon_{n,s}.\end{aligned} \tag{6.2a}$$

Moreover, the expected value and variance of the actual deviations from schedule of bus n at that stop are:

$$\begin{aligned}\mathbb{E}[\varepsilon_{n,s+1}] &= \mathbb{E}[\mathbb{E}[\varepsilon_{n,s+1} \mid \varepsilon_{n,s}]] \\ &= \mathbb{E}[\mathbb{E}[f_0\varepsilon_{n,s} + \nu_{n,s+1}]] \\ &= \mathbb{E}[f_0\varepsilon_{n,s}] = 0.\end{aligned} \tag{6.2b}$$

$$\begin{aligned}\text{var}[\varepsilon_{n,s+1}] &= \text{var}[\mathbb{E}[\varepsilon_{n,s+1} \mid \varepsilon_{n,s}]] + \mathbb{E}[\text{var}[\varepsilon_{n,s+1} \mid \varepsilon_{n,s}]] \\ &= \text{var}[f_0\varepsilon_{n,s}] + \mathbb{E}[\sigma^2] \\ &= f_0^2\sigma_\varepsilon^2 + \sigma^2 = \sigma_\varepsilon^2.\end{aligned} \tag{6.2c}$$

Note that the total expected value and the variance of $\varepsilon_{n,s+1}$ after suffering a GPS malfunction between stations s and $s+1$, does not vary from the steady state value. The reason is simple, the GPS malfunction occurred after holding was applied at station s, and the next holding action is applied after the bus arrival to $s+1$. Unfortunately, as the bus proceeds from this stop and the estimated deviations $\hat{\varepsilon}_{n,\cdot}$ are used in the holding maneuver, these values diverge from the steady state. If the simple control is applied at station $s+1$ using $\hat{\varepsilon}_{n,s+1}$ instead of the real $\varepsilon_{n,s+1}$, the state equation is:

$$\varepsilon_{n,s+2} = (1+\beta)\varepsilon_{n,s+1} + (f_0 - 1 - \beta)\hat{\varepsilon}_{n,s+1} + \nu_{s+2}. \tag{6.2d}$$

Therefore, the variance of the deviation from schedule for bus n at station $s+2$ would be:

$$\begin{aligned}\text{var}[\varepsilon_{n,s+2}] &= (1+\beta)^2\text{var}[\varepsilon_{n,s+1}] + (f_0 - 1 - \beta)^2\text{var}[\hat{\varepsilon}_{n,s+1}] + \text{var}[\nu_{s+2}] \\ &= (1+\beta)^2\sigma_\varepsilon^2 + (f_0 - 1 - \beta)^2 f_0^2\sigma_\varepsilon^2 + \sigma^2.\end{aligned} \tag{6.2e}$$

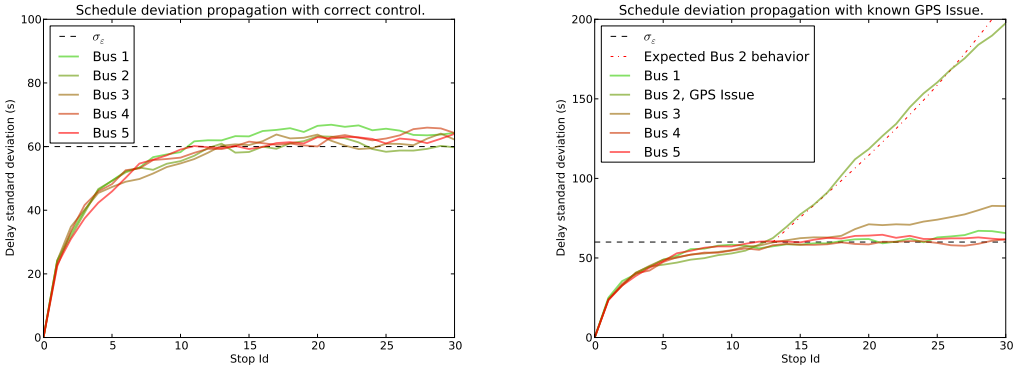
Similarly, for other stops along the route the schedule deviation variance could be iteratively computed as:

$$\begin{aligned}\text{var}[\varepsilon_{n,s+j}] &= (1+\beta)^2\text{var}[\varepsilon_{n,s+j-1}] + (f_0 - 1 - \beta)^2\text{var}[\hat{\varepsilon}_{n,s+j-1}] + \text{var}[\nu_{s+j}] \\ &= (1+\beta)^2\text{var}[\varepsilon_{n,s+j-1}] + (f_0 - 1 - \beta)^2 f_0^{2(j-1)}\sigma_\varepsilon^2 + \sigma^2\end{aligned} \tag{6.2f}$$

Unfortunately, this turns out to not be much better than the results for no control on bus n after the GPS issue arises. To see this, assume that $D_{n,s} = d_s$ at all stops in the affected bus and follow the previous calculations. The reader can verify that in an uncontrolled setting, the propagation of the deviations from schedule for the affected bus would have the following form:

$$\text{var}[\varepsilon_{n,s+j}] = (1+\beta)^2\text{var}[\varepsilon_{n,s+j}] + \beta^2\sigma_\varepsilon^2 + \sigma^2. \tag{6.3}$$

Note that the recursive term has exactly the same multiplying factor in both equations (6.2f) and (6.3). Thus, both sequences will have equal profiles as the number of visited stops increases.



(a) Correct control scenario.

(b) GPS malfunction, control based on estimated deviation from schedule.

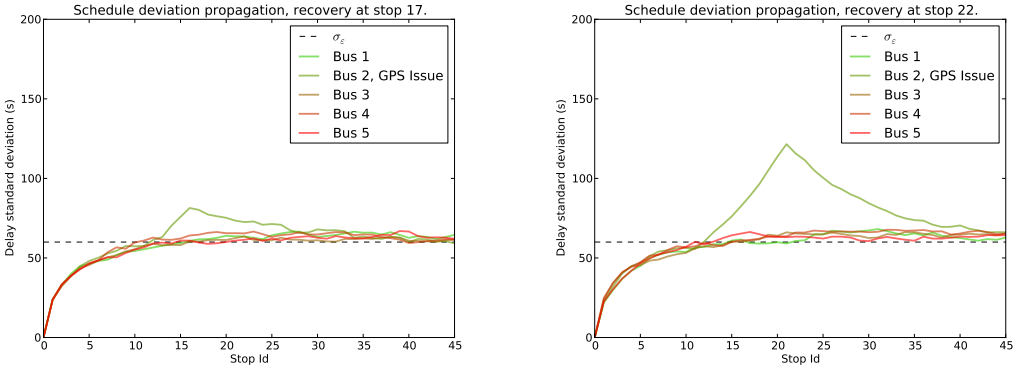
Figure 6.9: Average simulated $\sigma_\varepsilon(\boldsymbol{f}, n, s)$ in a homogeneous bus line with 5 buses.

Simulation Analysis A simulation is now used to illustrate the effects of a GPS outage. The goal is to compare the standard deviation of the deviations from schedule for buses in a controlled scenario versus a scenario where a GPS malfunction occurs. The simulation tracks a platoon of five buses in an idealized homogeneous line with an average cruising speed of 20 km/hr, a system headway of $H = 5$ minutes, a homogeneous demand parameter $\beta = 0.05$, an average boarding time per passenger $t_b = 3$ seconds, homogeneous noise with a standard deviation $\sigma = 24.7$ seconds, desired standard deviation of the deviations from schedule $\sigma_\varepsilon = 60$ seconds, optimal slack time $d^* = 26.53$s and optimal simple control coefficient $f_0^* = 0.9113$.

The buses are deterministically dispatched on time from a terminal station. The motion of the buses is simulated 500 times for each one of the two scenarios while they cruise between a sequence of 30 stations, without including the terminal. The simulated GPS issue arises in the second bus of the platoon while it cruises between stations $s = 11$ and $s = 12$, since it takes approximately 10 stations for these buses to reach their steady state. Figure 6.9 shows the averages of the observed $\sigma_\varepsilon(\boldsymbol{f}, n, s)$ across 500 simulation runs of the two scenarios previously described.

As we anticipated at the beginning of this section, this figure reveals the robustness of the simple control. Even though it is not possible to perfectly control the affected bus, the distributed nature of the control keeps the other buses in sync. In addition to the affected bus, only the bus following it slowly deviates from the desired σ_ε value over the next 18 stations after the issue arises. This is an encouraging result since it shows that the control method is resilient to isolated issues in the system.

But how well would the system recover from such a disruption? Two new batteries of simulations answer this question by following the motion of the previous 5 buses over 45 stations. In the first one, the affected bus recovers its GPS 5 stops after suffering the



(a) GPS recovery after 5 stations.

(b) GPS recovery after 10 stations.

Figure 6.10: Evolution of the average simulated $\sigma_\varepsilon(\boldsymbol{f}, n, s)$ after GPS recovery at two different stations.

problem, at stop $s = 17$. In the second, the recovery occurs 10 stops after the disruption to the GPS service arises. In both cases, once the GPS is recovered, the original simple control is immediately applied to the affected bus without any changes to the slack or the control coefficient f_0^*. The results of these simulations are shown in Figure 6.10. In both cases, one can see how the values of $\sigma_\varepsilon(\boldsymbol{f}, n, s)$ transition back to the desired $\sigma_\varepsilon = 60$ seconds during its trip through the simulated 45 stations.

These results show that by using the proposed dynamic holding control methodology, isolated problems do not destabilize the entire system. Instead, the issue is confined to the affected bus and its immediate follower. Moreover, recovery is possible without any special actions, by simply restoring the original control when the GPS signal returns. Thus, if the system is correctly monitored, something which is possible with existing technology, sporadic problems should not pose a significant threat to overall performance.

Chapter 7

Contributions and Future Work

This section summarizes the core results of this dissertation. It also identifies avenues that could be explored in the future within the transit operations and control arena.

7.1 Main Contributions

This dissertation attempted to extend the applicability range of dynamic holding methods for bus systems. To do so, a fairly general dynamic holding control framework was formulated using the notion of a virtual schedule. This framework encompasses both schedule and headway-based operations as special cases. It was shown that unlike existing methods, the proposed form of dynamic holding control could not only guarantee headway regularity, but also maintain schedule adherence in isolated lines. This is a fundamental outcome since most agencies operate their bus lines under a planned schedule and therefore can benefit directly from this result. The proposed form of control feeds frequent but small corrective measures to the drivers before problems are severe. This improves service reliability and overcomes bunching without decreasing the line's commercial speed too much.

The analysis was extended to show that a version of the general linear dynamic holding control, called "simple control", could achieve quasi-optimal results. The simple control only requires information about the arriving bus and the preceding bus at a given station to compute the holding times. The need for a small amount of information and the distributed nature of the control facilitate its application in real systems. The communication and hardware requirements were shown to be minimal.

The previous results consider an idealized framework: an infinite line with homogeneous demand conditions. Even though it was been argued that these assumptions define a worst case scenario for bus control, the fact is that such conditions are not realistic and may be too pessimistic for a real world scenario. For this reason, the modeling framework was reformulated to capture the nature of a looping line with location dependent demand. An optimization problem that allows the control parameters to be chosen was also formulated.

The proposed control approach in a heterogeneous system was tested using data collected

from a real-world system (Bear Transit at U.C. Berkeley). A discrete event-based simulation tool that emulates the operation of a bus line was developed. The simulation was calibrated using various operational parameters measured in the field such as travel time variation and demand levels. The results obtained after optimizing different families of control methods (simple control, headway-based controls, etc.) and generating multiple batteries of simulations qualitatively matched the theoretical predictions assessing homogeneous features.

The control formulation was extended to cover corridors where multiple bus lines overlap. First, the models that characterize the motion of the buses in the system were adapted to this extended scenario. A general dynamic holding control was proposed. It was shown that a particular form of this control could make the different bus lines in the corridor behave as if they were operating independently from one another. As with the isolated line, the control could also bound both schedule and headway deviations.

The simulation tool was further improved to test the theory for corridors with overlapping lines. In this case, the data were collected for a corridor including two lines in San Sebastián, Spain, using on-board devices that monitored the system for various months. The simulations predicted significant gains in reliability without affecting the underlying commercial speed. The proposed control was then implemented in the corridor, using on-board tablets to communicate the control decisions to the drivers. The before and after results were then compiled using field measurements and they were compared to the simulation results. Significant improvements were also observed in the field implementation.

The last part of the dissertation identified problems that appear in real world applications which can negatively affect the performance of the control. These issues are generally ignored in the literature and this section sheds light on their effects. In particular, GPS malfunctions (the most serious of the problems) were analyzed, showing the capability of the control to recover from temporary disruptions.

7.2 Future Work

The work done so far can be seen as the tip of the iceberg. There are multiple directions in which this research could be expanded. The following subsections highlight some of them.

7.2.1 Non-linear and Heuristic Control Strategies

Transit dispatchers sometimes resort to fairly radical forms of control to try to revert buses with large delays back to their schedule. Some of this strategies include skipping multiple stops, negating passenger boardings or cutting the final stretch of a particular trip. These forms of intervention are generally applied in an ad-hoc manner.

In the event of large disruptions, it is possible that system recovery with the application of holding could be slow and, in some cases, almost not possible. These alternative techniques could be used to truly accelerate the motion of the buses and revert them back to their schedule in a faster way. For this reason, it seems appropriate to analyze the application

of dynamic holding control in conjunction with other control techniques. Performing an analytic assessment of this sort of approach may prove difficult but the developed simulation could certainly help.

7.2.2 Operational Possibilities: Dynamic Transfer Coordination

This work has analyzed dynamic holding control methods as a mean to improve the quality of service in one or multiple lines. However, the possibility of using the proposed control to facilitate transfers between different bus lines or between buses and other transportation modes remains unexplored. With the proposed formulation it is possible to obtain a good picture of the variability in the arrival times of buses at a particular location. This information could be used to probabilistically assess the wait times that transferring passengers will face. This type of work would be highly relevant for those type of systems where transfers are an integral part of the average trip, see; [25, 32].

7.2.3 Assessing Human Factors

The success of any form of control depends entirely on the willingness of the drivers to follow the recommendations. Without drivers' cooperation the system is doomed to fail. For this reason, it is essential to understand what are the factors that influence bus drivers in their decision-making process. A potential avenue of study could focus on determining which operational and environmental conditions would favor driver compliance. The tools to do so are already in place since the on-board units used in our case study allow a system manager to monitor driving tasks with a significant level of precision.

7.2.4 Interface design

Another area that would require further study is the design of the control interface. In the San Sebastián case study, the drivers were directly involved in the design process, which resulted in better acceptance. However, the design could probably benefit from a more systematic approach. Human-machine interface design techniques could be applied to fine tune the on-board unit interface and improve driver compliance levels.

Finally, the interface could be further transformed by undertaking a gamification approach. The idea is to use game mechanics in the context of the driving tasks with suitable rewards. A possible goal could be the improvement of driver compliance with the control recommendations. This sort of technique has shown promising results in several fields [29] and transit systems seem ripe to benefit from it.

Bibliography

[1] M. Abkowitz, I. Eiger, and I. Engelstein. "Optimal control of headway variation on transit routes". In: *Journal of Advanced Transportation* 20.1 (1986), pp. 73–88.

[2] M. Abkowitz and I. Engelstein. "Method for maintaining transit service regularity". In: *Transportation Research Record* 961 (1984), pp. 1–8.

[3] A. Adamski and A. Turnau. "Simulation support tool for real-time dispatching control in public transport". In: *Transportation Research Part A* 32.2 (1998), pp. 73–87.

[4] VIA Analytics. *VIZ Metrics Demo, a platform to explore transit data.* 2014. URL: `https://viz.v-a.io/demo/metrics` (visited on 10/28/2014).

[5] J. Argote et al. "An Analysis of the Bear Transit System Operations". A report prepared for the University of California at Berkeley Department of Parking and Transportation. 2011.

[6] Daniel Ashbrook and Thad Starner. "Using GPS to learn significant locations and predict movement across multiple users". In: *Personal and Ubiquitous Computing* 7.5 (2003), pp. 275–286.

[7] Kittelson & Associates et al. *Transit Capacity and Quality of Service Manual*. Vol. 100. Transportation Research Board, 2003.

[8] Chicago Transit Authority. *Chicago Transit Authority Performance Metrics Reports.* 2012. URL: `http://www.transitchicago.com/perfmetre.aspx`.

[9] M. Baaj and H. Mahmassani. "Hybrid route generation heuristic algorithm for the design of transit networks". In: *Transportation Research Part C* 3.1 (1995), pp. 31–50.

[10] A. Barnett. "On controlling randomness in transit operations". In: *Transportation Science* 8.2 (1974), pp. 101–116.

[11] John J Bartholdi III and Donald D Eisenstein. "A self-coördinating bus route to resist bus bunching". In: *Transportation Research Part B: Methodological* 46.4 (2012), pp. 481–491.

[12] J. W. Bates. "Definition of Practices for Bus Transit On-Time Performance: Preliminary Study." In: *Transportation Research Circular* 300 (1986), pp. 1–5.

[13] J. Berechman. "Costs, Economies of Scale and Factor Demand in Bus Transport: An Analysis". In: *Journal of Transport Economics and Policy* 17.1 (1983), pp. 7–24.

[14] Robert L Bertini and Sutti Tantiyanugulchai. "Transit buses as traffic probes: Use of geolocation data for empirical evaluation". In: *Transportation Research Record: Journal of the Transportation Research Board* 1870.1 (2004), pp. 35–45.

[15] Joseph-Frédéric Bonnans et al. "Numerical optimization: theoretical and practical aspects". In: Local Convergence: BFGS. Springer, 2006. Chap. 4 Newtonian Methods.

[16] Larry A Bowman and Mark A Turnquist. "Service frequency, schedule reliability and passenger wait times at transit stops". In: *Transportation Research Part A: General* 15.6 (1981), pp. 465–471.

[17] Daniel K Boyle. "Passenger Counting Technologies and Procedures. TCRP Synthesis 29". In: *Transportation Research Board, National Academy Press, Washington, DC* (1998).

[18] A. Carrel, A. Halvorsen, and J.L. Walker. "Passengers' perception of and behavioral adaptation to unreliability in public transportation". In: *Transportation Research Board 92nd Annual Meeting*. 13-5289. 2013.

[19] M. V. Chester. "Life-cycle Environmental Inventory of Passenger Transportation in the United States". PhD thesis. University of California, Berkeley, 2008.

[20] Morgan T Conlon et al. "Successful arterial street limited-stop express bus service in Chicago". In: *Transportation Research Record: Journal of the Transportation Research Board* 1760.1 (2001), pp. 74–80.

[21] Denis Cousineau, Scott Brown, and Andrew Heathcote. "Fitting distributions using maximum likelihood: Methods and packages". In: *Behavior Research Methods, Instruments, & Computers* 36.4 (2004), pp. 742–756.

[22] C. Daganzo. *Fundamentals of Transportation and Traffic Operations*. Pergamond Oxford, 1997.

[23] C. F. Daganzo. "A headway-based approach to eliminate bus bunching: Systematic analysis and comparisons". In: *Transportation Research Part B* 43.10 (2009), pp. 913–921.

[24] C. F. Daganzo. "On the variational theory of traffic flow: well-posedness, duality and applications." In: *Networks and Heterogeneous Media* 1.4 (2006), pp. 601–619.

[25] C. F. Daganzo. "Structure of competitive transit networks". In: *Transportation Research Part B* 44.4 (2010), pp. 434–446.

[26] C. F. Daganzo and J. Pilachowski. "Reducing bunching with bus-to-bus cooperation". In: *Transportation Research Part B* 45.1 (2011), pp. 267–277.

[27] F. Delgado et al. "Real-time control of buses in a transit corridor based on vehicle holding and boarding limits." In: *Transportation Research Record* 2090 (2009), pp. 59–67.

[28] Maged Dessouky et al. "Bus dispatching at timed transfer transit stations using bus tracking technology". In: *Transportation Research Part C: Emerging Technologies* 7.4 (1999), pp. 187–208.

[29] Sebastian Deterding et al. "From game design elements to gamefulness: defining gamification". In: *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*. ACM. 2011, pp. 9–15.

[30] X. J. Eberlein, N. H. M. Wilson, and D. Bernstein. "The holding problem with real-time information available". In: *Transportation Science* 35.1 (2001), pp. 1–18.

[31] X.J. Eberlein. "Real time control strategies in transit operations: models and analysis." PhD thesis. Cambridge: Massachusetts Institute of Technology, 1995.

[32] M Estrada et al. "Design and implementation of efficient transit networks: Procedure, case study and validity test". In: *Transportation Research Part A: Policy and Practice* 45.9 (2011), pp. 935–950.

[33] Patrick Th Eugster et al. "The many faces of publish/subscribe". In: *ACM Computing Surveys (CSUR)* 35.2 (2003), pp. 114–131.

[34] European Commission. *White Paper on transport: Roadmap to a Single European Transport Area – Towards a Competitive and Resource-Efficient Transport System*. Tech. rep. Brussels: European Commission, Department of Mobility and Transport, 2011.

[35] L. Fu, Q. Liu, and P. Calamai. "Real-time optimization model for dynamic scheduling of transit operations". In: *Transportation Research Record* 1857 (2003), pp. 48–55.

[36] Liping Fu and Xuhui Yang. "Design and implementation of bus-holding control strategies with real-time information". In: *Transportation Research Record: Journal of the Transportation Research Board* 1791.1 (2002), pp. 6–12.

[37] P. G. Furth and F. B. Day. "Transit routing and scheduling strategies for heavy demand corridors". In: *Transportation Research Record* 1011 (1985), pp. 23–26.

[38] Peter Gregory Furth et al. *Uses of archived AVL-APC data to improve transit performance and management: Review and potential*. Transportation Research Board Washington, DC, 2003.

[39] T. F. Golob et al. "An analysis of consumer preferences for a public transportation system". In: *Transportation Research* 6.1 (1972), pp. 81–102.

[40] Google. *General Transit Feed Specification (GTFS)*. 2014. URL: `https://developers.google.com/transit/gtfs/` (visited on 09/30/2013).

[41] W. Grega. "A decomposition algorithm for the determination of optimal bus frequencies". In: *System Modelling and Optimization* 84 (1986), pp. 292–297.

[42] W. Gu et al. "On the capacity of isolated, curbside bus stops". In: *Transportation Research Part B* 45.4 (2011), pp. 714–723.

[43] Zhan Guo and Nigel HM Wilson. "Assessment of the Transfer Penalty for Transit Trips Geographic Information System-Based Disaggregate Modeling Approach". In: *Transportation Research Record: Journal of the Transportation Research Board* 1872.1 (2004), pp. 10–18.

[44] A. F. Han and N. H. M. Wilson. "The allocation of buses in heavily utilized networks with overlapping routes". In: *Transportation Research Part B* 16.3 (1982), pp. 221–232.

[45] M. D. Hickman. "An analytic stochastic model for the transit vehicle holding problem". In: *Transportation Science* 35.3 (2001), pp. 215–237.

[46] Mark Hickman. "Bus automatic vehicle location (AVL) systems". In: *Assessing the Benefits and Costs of ITS*. Springer, 2004, pp. 59–88.

[47] D. Koffman. "A simulation study of alternative real-time bus headway control strategies". In: *Transportation Research Record* 663 (1978), pp. 41–46.

[48] Zvi Kohavi and Niraj K Jha. *Switching and finite automata theory*. Cambridge University Press, 2010.

[49] Wen-Tai Lai and Ching-Fu Chen. "Behavioral intentions of public transit passengers-The roles of service quality, perceived value, satisfaction and involvement". In: *Transport Policy* 18.2 (2011), pp. 318–325.

[50] Der-Horng Lee, Lijun Sun, and Alex Erath. "Study of bus service reliability in singapore using fare card data". In: *12th Asia-Pacific Intelligent Transpotation Forum*. 2012.

[51] C. Leiva et al. "Design of limited-stop services for an urban bus corridor with capacity constraints". In: *Transportation Research Part B* 44.10 (2010), pp. 1186–1201.

[52] Y. Li, J. Rousseau, and M. Gendreau. "Real-time dispatching public transit operations with and without bus location information". In: *Computer-aided scheduling* 430 (1993), pp. 296–308.

[53] Jeffrey Lidicker et al. "Shuttle Transit System Evaluation Methodology: Performance, Characterization, and Optimization". In: *Transportation Research Board 93rd Annual Meeting*. 14-3481. 2014.

[54] Pedro Lizana et al. "Bus Control Strategy Application: Case Study of Santiago Transit System". In: *Procedia Computer Science* 32 (2014), pp. 397–404.

[55] Yuko J Nakanishi. "PART 1: Bus: Bus Performance Indicators: On-Time Performance and Service Regularity". In: *Transportation Research Record: Journal of the Transportation Research Board* 1571.1 (1997), pp. 1–13.

[56] Dmitry Namiot and Manfred Sneps-Sneppe. "Geofence and Network Proximity". In: *Internet of Things, Smart Spaces, and Next Generation Networking*. Springer, 2013, pp. 117–127.

[57] G. F. Newell and R. B. Potts. "Maintaining a bus schedule". In: *Proceedings of the 2nd Australian Road Research Board*. Vol. 2. 1964, pp. 388–393.

[58] G.F. Newell. "Control of pairing of vehicles on a public transportation route, two vehicles, one control point". In: *Transportation Science* 8.3 (1974), pp. 248–264.

[59] J. de D. Ortuzar. "Nested Logit Models for Mixed-Mode Travel in Urban Corridors". In: *Transportation Research Part A* 17.4 (1983), pp. 283–299.

[60] E. E. Osuna and G. F. Newell. "Control strategies for an idealized public transportation system". In: *Transportation Science* 6.1 (1972), pp. 52–72.

[61] S. B. Pattnaik, S. Mohan, and V. M. Tom. "Urban bus transit route network design using genetic algorithm". In: *Journal of Transportation Engineering* 124.4 (1998), pp. 368–375.

[62] Marie-Pier Pelletier, Martin Trépanier, and Catherine Morency. "Smart card data use in public transit: A literature review". In: *Transportation Research Part C: Emerging Technologies* 19.4 (2011), pp. 557–568.

[63] R Tyrrell Rockafellar. "Convex analysis". In: 28. Princeton university press, 1997. Chap. Section 32: the Maximum of a Convex Function.

[64] Mark Roseland. *Toward sustainable communities: solutions for citizens and their governments*. New Society Publishers, 2013.

[65] David Schrank, Bill Eisele, and Tim Lomax. *TTIs 2012 Urban Mobility Report*. Tech. rep. 2012.

[66] P. N. Senevirante. "Analysis of on-time performance of bus services using simulation". In: *Transportation Engineering* 116.4 (1990), pp. 517–531.

[67] Bernard W Silverman. *Density estimation for statistics and data analysis*. Vol. 26. CRC press, 1986.

[68] W. Suh, K. Chon, and S. Rhee. "Effect of skip-stop policy on a korean subway system". In: *Transportation Research Record* 1793 (2002), pp. 33–39.

[69] A. Sun and M. Hickman. "The holding problem at multiple holding stations". In: *Computer-aided systems in public transport* 600.3 (2008), pp. 339–359.

[70] A. Sun and M. Hickman. "The real-time stop skipping problem". In: *Transportation Systems* 9.2 (2005), pp. 91–109.

[71] Arvind Thiagarajan et al. "Cooperative transit tracking using smart-phones". In: *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*. ACM. 2010, pp. 85–98.

[72] Mark A Turnquist and Steven W Blume. "Evaluating potential effectiveness of headway control strategies for transit systems". In: *Transportation Research Record* 746 (1980).

[73] Yannis Tyrinopoulos and Constantinos Antoniou. "Public transit user satisfaction: Variability and policy implications". In: *Transport Policy* 15.4 (2008), pp. 260–272.

[74] U. Vandebona and A. Richardson. "Effect of checkpoint control strategies ina simulated transit operation". In: *Transportation Research Part A* 20.6 (1986), pp. 429–436.

[75] R. J. Wallin and P. H. Wright. "Factors which influence modal choice". In: *Traffic Quarterly* 28.2 (1974), pp. 271–289.

[76] Mark Wardman. "Public transport values of time". In: *Transport policy* 11.4 (2004), pp. 363–377.

[77] Y. Xuan, J. Argote, and C. Daganzo. "Dynamic Bus Holding Strategies for Schedule Reliability: Optimal Linear Control and Performance Analysis". In: *Transportation Research Part B* 45 (2011), pp. 1831–1845.

Appendix A

Gradients for the greedy search

In developing Figure 3.1, we solved the following optimization problem, which is equivalent to (MP3.2):

(MPA.1)
$$\begin{aligned} &\min_{\boldsymbol{f}} \sigma_D^2(\boldsymbol{f}) \\ &\text{s.t. } \sigma_\varepsilon^2(\boldsymbol{f}) \leq s_\varepsilon^2. \end{aligned}$$

From equations (3.11b) and (3.14b), we find the following expressions for $\sigma_\varepsilon^2(\boldsymbol{f})$ and $\sigma_D^2(\boldsymbol{f})$, whose gradients we seek:

$$\sigma_\varepsilon^2(\boldsymbol{f}) = \sigma^2 \sum_{j=0}^{\infty} \sum_i (f_{i|j})^2, \tag{A.1}$$

$$\sigma_D^2(\boldsymbol{f}) = \sigma^2 \sum_{j=0}^{\infty} \sum_i [(1+\beta) f_{i|j} - \beta f_{i-1|j} - f_{i|j+1}]^2. \tag{A.2}$$

Using generating functions, one can find the following expression for $f_{i|j}$:

$$f_{i|j} = \sum_{\substack{k_0,k_1,\cdots,k_{m-1} \\ \sum_{r=0}^{m-1} k_r = j \\ \sum_{r=0}^{m-1} r k_r = i}} \frac{j!}{k_0!\, k_1! \cdots k_{m-1}!} f_0^{k_0} f_1^{k_1} \cdots f_{m-1}^{k_{m-1}}. \tag{A.3}$$

Since

$$\frac{\partial f_{i|j}}{\partial f_r} = j f_{i-r|j-1}, \tag{A.4}$$

the partial derivatives of $\sigma_\varepsilon^2(\boldsymbol{f})$ and $\sigma_D^2(\boldsymbol{f})$ with respect to f_r can be expressed as:

$$\frac{\partial \sigma_\varepsilon^2}{\partial f_r} = 2\sigma^2 \sum_{j=0}^{\infty} \sum_i (f_{i|j})(j f_{i-r|j-1}), \tag{A.5}$$

and

$$\frac{\partial \sigma_D^2}{\partial f_r} = 2\sigma^2 \sum_{j=0}^{\infty} \sum_{i} [(1+\beta) f_{i|j} - \beta f_{i-1|j} - f_{i|j+1}] \qquad \text{(A.6)}$$
$$[(1+\beta) j f_{i-r|j-1} - \beta j f_{i-r-1|j-1} - (j+1) f_{i-r|j}].$$

Appendix B

Locating Control Points

We have assumed during the analysis that the proposed control method is applied at each station, but this is not always desirable. As shown in [22, 23], it is often beneficial to space out the control points more widely. With this in mind, it is assumed here that control points are located every K stations, and K is treated as a decision variable. The demand rate β is assumed to be uniform throughout the bus line and such that $\beta << 1$.

We will first transform the equations of bus motion from station to station into approximately equivalent equations describing the bus motion from control point to control point, as if there were no intermediate stations. It will be shown that when $\beta << 1$, one may simply replace β and σ with $\beta' = K\beta$ and $\sigma' = \sigma\sqrt{K + K(K-1)\beta}$ to model the bus motion in this manner.

The variance of the noise between consecutive control points $(\sigma')^2$ is simply given by (3.11a) using the $f_{i|j}$ of the uncontrolled case and replacing s by K. Recall that for the uncontrolled situation, $\boldsymbol{f} = [\cdots, 0, 1+\beta, -\beta, 0, \cdots]^T$. It can be seen from the binomial expansion that

$$f_{i|j} = \binom{i}{m}(1+\beta)^{j-i}(-\beta)^i. \tag{B.1}$$

Thus,

$$\begin{aligned}\boldsymbol{f}_{|j} &= [\cdots, (1+\beta)^j, j(1+\beta)^{j-1}(-\beta), \binom{j}{2}(1+\beta)^{j-2}(-\beta)^2, \cdots]^T \\ &= [\cdots, 1+j\beta, -j\beta, \cdots]^T.\end{aligned} \tag{B.2}$$

The last approximation works because $\beta << 1$ and thus we can neglect terms of order two and higher, $\beta^i \approx 0$, $i \geq 2$. Equation (B.2) can now be inserted in (3.11a) to yield:

$$
\begin{aligned}
(\sigma')^2 &= \sigma^2 \sum_{j=0}^{K-1} \sum_{i} (f_{i|j})^2 \\
&= \sigma^2 \sum_{j=0}^{K-1} \sum_{i} [\binom{j}{i} (1+\beta)^{j-i} (-\beta)^i]^2 \\
&\approx \sigma^2 \sum_{j=0}^{K-1} (1 + 2j\beta) \\
&= \sigma^2 (K + K(K-1)\beta). \qquad \text{(B.3)}
\end{aligned}
$$

It should also be clear from (B.2) that when $j = K$ and $\beta << 1$, the dimensionless demand between control points is $\beta' = K\beta$. By setting $\beta' = K\beta$ and $\sigma' = \sigma\sqrt{K + K(K-1)\beta}$, we can treat the bus motion as if there were only stations at the control points and apply (MP3.2) with these new parameters. The number of stations between control points, K, is however, a decision variable in the new version of (MP3.2). Consideration of (3.11b) and (3.14b) reveals that this new mathematical program is:

(MPA.1)
$$
\begin{aligned}
\min_{\boldsymbol{f},K} \; & \frac{\sigma'}{\sigma} \frac{d(\boldsymbol{f}, K\beta)}{K} \\
\text{s.t.} \; & \frac{\sigma'}{\sigma} \sigma_\varepsilon(\boldsymbol{f}) \leq s_\varepsilon \\
& \frac{\sigma'}{\sigma} = \sqrt{K + K(K-1)\beta}.
\end{aligned}
$$

This MP can be solved numerically as in Subsection 3.3.1, or analytically if we adopt the simple control method. A similar program can be written if the objective is to guarantee a headway variance by using σ_h instead of σ_ε.

In some instances, it may be useful to have an idea of the average of the variances of the schedule deviations across all the stations, $\bar{\sigma}_\varepsilon^2$; or the average of the headway variances, $\bar{\sigma}_h^2$. To obtain these values, (3.11b) and (3.13b) would still be used, but the recursion $\boldsymbol{f}_{|j} = \boldsymbol{f} * \boldsymbol{f}_{|j-1}$ used to evaluate the convolution coefficients, $f_{i|j}$, would use the f_i from (3.4c) when j corresponds to a control point, and $f_0 = 1 + \beta_{n,s}$, $f_1 = \beta_{n,s}$ and $f_i = 0\, i \notin 0, 1$ from (3.5) otherwise.

Finally, note that this equivalent demand and noise parameters, β' and σ', will be relevant to characterize the motion of buses in segments of routes where control cannot be applied. This would be the case in an area where GPS signal is not available and no other tracking methods are in place (e.g. bluetooth beacons, triangulation using WiFi signal, etc.). In those cases, β' and σ' should be calculated considering those stops located within the area affected by the GPS signal shortage.

Appendix C

Bear Transit Data Collection.

The data collection effort to characterize the Bear Transit system consisted on two different parts. First, surveys handed to on-board riders were used to infer the system's origin/destination matrix. These surveys allowed to capture the boarding stop and intended alighting stop for approximately 95% of the riders during a period of one week. More information on the data collection effort can be found in [53]. The resulting origin/destination used in Chapter 4 analysis can be found in table C.1.

Table C.1: Perimeter line origin destination matrix values. A cell in row i and column j provides the probability that a passenger boarding at station i will alight at station j.

		Destinations														
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Origins	0	0.00	0.01	0.02	0.13	0.20	0.46	0.07	0.03	0.05	0.01	0.01	0.00	0.00	0.00	0.00
	1	0.00	0.00	0.03	0.11	0.24	0.29	0.06	0.09	0.10	0.01	0.01	0.04	0.00	0.00	0.01
	2	0.00	0.00	0.00	0.01	0.23	0.15	0.07	0.07	0.18	0.15	0.04	0.08	0.03	0.00	0.00
	3	0.13	0.04	0.00	0.04	0.09	0.04	0.00	0.13	0.26	0.13	0.00	0.04	0.04	0.00	0.04
	4	0.18	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.18	0.08	0.10	0.23	0.05	0.03	0.15
	5	0.19	0.03	0.02	0.01	0.00	0.01	0.01	0.00	0.17	0.09	0.12	0.07	0.13	0.11	0.04
	6	0.60	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.00	0.20	0.00	0.00	0.00
	7	0.38	0.13	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.00	0.00	0.25
	8	0.21	0.00	0.14	0.14	0.07	0.18	0.00	0.00	0.00	0.00	0.00	0.07	0.04	0.04	0.11
	9	0.25	0.07	0.04	0.02	0.04	0.05	0.00	0.00	0.00	0.00	0.00	0.13	0.11	0.13	0.18
	10	0.35	0.09	0.03	0.06	0.03	0.09	0.00	0.00	0.00	0.00	0.00	0.09	0.09	0.12	0.06
	11	0.17	0.06	0.02	0.04	0.17	0.14	0.02	0.04	0.00	0.01	0.00	0.00	0.05	0.10	0.17
	12	0.21	0.06	0.03	0.12	0.18	0.12	0.06	0.03	0.12	0.00	0.03	0.06	0.00	0.00	0.00
	13	0.04	0.00	0.04	0.12	0.08	0.24	0.16	0.04	0.12	0.00	0.04	0.08	0.04	0.00	0.00
	14	0.00	0.00	0.08	0.13	0.25	0.33	0.08	0.08	0.04	0.00	0.00	0.00	0.00	0.00	0.00

Second, cruising times between stations and boardings/alightings rates at stations were obtained using a semi-automatic data collection tool. Figure C.1 shows a screenshot of the software's user interface. The tool is a simple MATLAB script that allows an on board auditor to quickly log a set of pre-specified events: skipped stop, door opening, door closing,

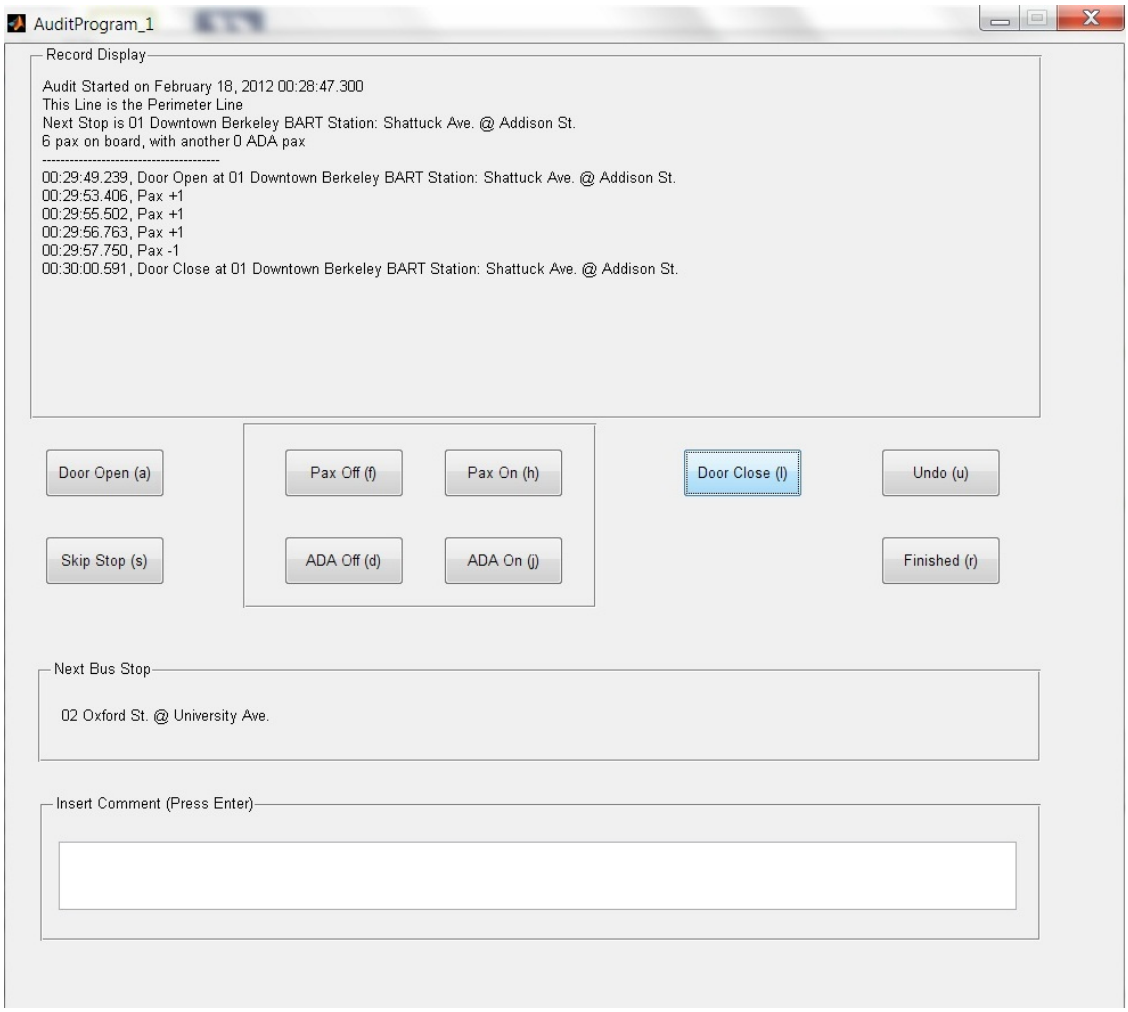


Figure C.1: User interface of the data collection tool used to audit the Bear Transit system.

regular passenger boarding, regular passenger alighting, ADA passenger boarding, ADA passenger alighting.

The logs captured by the auditors were combined offline into a single file. That file was first manually parsed to avoid any inconsistencies[1]. The resulting filtered file was automatically processed obtaining the sought average parameters on a per hour basis. Table C.2 shows the Perimeter line parameters, including demand rates at each station and the mean and standard deviation of link travel times. The results were averaged over the 10am to 12am period of time during Mondays through Fridays. This is the time interval simulated in Chapter 4 to compare the multiple linear holding strategies.

[1]Auditors were allowed to insert comments as part of the logs. This feature was occasionally used by the auditors to make clarifications while conducting their logging efforts.

Table C.2: Average parameters for the Perimeter line, obtained on Mondays through Thursdays, 10am to 12am, April 2011.

Stop Index s	Station Shortname	Postmile (km)	$\beta_{\cdot,s}$	$c_{\cdot,s}$ (sec)	$\sigma_{\cdot,s}$ (sec)
0	BART	0	0.021	143.0	13.7
1	University & Oxford	0.31	0.007	104.7	11.9
2	Hearst & Arch	0.61	0.014	70.9	5.4
3	Hearst & Euclid	0.97	0.006	73.4	9.3
4	Hearst & Cory Hall	1.14	0.017	145.7	13.6
5	Hearst Mining Circle	1.79	0.017	102.8	13.0
6	Hall & Gayley	2.17	0.003	56.6	2.1
7	Haas Business School	2.33	0.003	44.6	2.2
8	International House	2.54	0.004	94.8	13.8
9	Bancroft & College	2.80	0.006	53.3	3.7
10	Bancroft & Bowditch	2.95	0.003	102.1	11.2
11	Bancroft & Telegraph	3.19	0.008	71.2	10.2
12	Bancroft & Ellsworth	3.38	0.003	80.4	6.1
13	Bancroft & Fulton	3.91	0.004	44.4	4.6
14	Shattuck & Kittredge	4.01	0.007	69.1	8.3

Appendix D

Multi-line Stability and Control Optimization

D.1 Stability Analysis

Since the proposed control method in a multi-line corridor yields state equations that share the same functional form as in an isolated line scenario the results obtained in Section 3.2 will hold once again. The main difference in this new scenario resides in the need to prove the stability of two different types of headways: intra-line headways between buses in the same line and intra-line headways between buses of distinct lines.

Schedule adherence

The stability of the deviations from schedule follows immediately from the results in Section 3.2 for the simplified canceling holding control method in equations (5.4d) and (5.6b). In both cases the motion equations that characterize the propagation of the deviations from schedule has the same form as in equation (3.10b):

$$\varepsilon^l_{n_l,s} = \sum_{j=0}^{s-1} \sum_i f^l_{i|j} \nu_{n_l-i,s-j}, \tag{D.1}$$

The only difference is that the unique control coefficients $\boldsymbol{f}$ in the isolated line scenario are substituted by the line-specific control coefficients $\boldsymbol{f}^l$ in the multi-line case. Thus, Theorem 3.2.2 establishing the condition for bounded deviations from schedule will also apply for this particular form of the dynamic holding control in a homogeneous corridor with multiple overlapping lines. As long as $\sum_i |f^l_i| < 1$ the deviations from schedule of buses in line l will be bounded.

Intra-line headway adherence

In Chapter 5 there was no reference to a pre-planned headway. However, it is possible to define the scheduled intra-line headway for bus n_l at station s as $H^l_{n_l,s} = t_{n_l,s} - t_{n_l-1,s}$. On the other hand, the actual intra-line headway could be defined as $h^l_{n_l,s} = a_{n_l,s} - a_{n_l-1,s} = H^l_{n_l,s} + \varepsilon_{n_l,s} - \varepsilon_{n_l-1,s}$. Thus, considering the equivalent of equation (3.10b) for the multi-line case, we have that the intra-line headway could be formulated as:

$$h^l_{n_l,s} = H^l_{n_l,s} + \sum_{j=0}^{s-1} \sum_{i} (f^l_{i|j} - f^l_{i-1|j}) \nu_{n_l-i,s-j}, \tag{D.2}$$

The resulting expression has the same form as equation (3.13a). Moreover, Corollary 3.2.6 would also hold, showing that any control method that would bound the deviations from schedule would also bound the intra-line headways.

Inter-line headway adherence

Let us assume that we are interested in the inter-line headway between buses from line l_1 and line l_2 at station s. Without loss of generality, consider that bus n_{l_1} from line l_1 was preceded by bus n_{l_2} at that station, so that $T_{L^m}(n_{l_1}) = n_{L^m_1}$, $T_{L^m}(n_{l_2}) = n_{L^m_2}$, and $n_{L^m_1} > n_{L^m_2}$. It is further assumed that $n_{L^m_1} = n_{L^m_2} + n_t$, which means that $n_t - 1$ buses from lines other than l_1 arrived at station s after bus n_{l_2} and before bus n_{l_1} did. Then the inter-line headway could be formulated as:

$$h^{l_1 l_2}_{n_{l_1},s} = H^{l_1 l_2}_{n_{l_1},s} + \sum_{j=0}^{s-1} \sum_{i} f^{l_1}_{i|j} \nu_{n_{l_1}-i,s-j} - \sum_{j=0}^{s-1} \sum_{i} f^{l_2}_{i|j} \nu_{n_{l_1}-i,s-j}, \tag{D.3}$$

In that case, Corollary 3.2.6 would hold, providing the sufficient conditions for bounded inter-line headway standard deviations in an infinitely long corridor with multiple overlapping lines.

D.2 Control Optimization

An essential part of the optimization of the control coefficients resides in the formulation of the slack times. To do so, one can follow the same approach as in Section 3.3, considering a formulation that probabilistically guarantees that the slack times $d^l_{n_l,s}$ never run short. Thus, the slack times are obtained as three times the variances of the holding times $D^l_{n_l,s}$. For the sake of brevity, let us focus on the two line scenario with the simple version of

the canceling dynamic holding control. Under these assumptions and using the convolution approach first introduced in Section 3.3 the holding times can be formulated as:

$$D^l_{n_l,s} = d^l_{n_l,s} - [(1+\beta^l_{n_l,s}+\beta^L_{n_L,s})\varepsilon_{n_l,s} - \beta^l_{n_l,s}\varepsilon_{n_l-1,s} - \beta^L_{n_L,s}\varepsilon_{n_L-1,s} - \sum_i f^l_i \varepsilon_{n_l-i,s}] \quad \text{(D.4a)}$$

$$= d^l_{n_l,s} - \sum_{j=0}^{s-1}\sum_i [(1+\beta^l_{n_l,s}+\beta^L_{n_L,s})f^l_{i|j} - \beta^l_{n_l,s}f^l_{i-1|j} - f^l_{i|j+1}]\nu_{n_l-i,s-j}$$

$$+ \sum_{j=0}^{s-1}\sum_{i'} [\beta^L_{n_L,s} f^{R_L(n_L-1)}_{i'|j}]\nu_{T_L^{-1}(n_L-1)-i',s-j}.$$

Where R_L is a function that converts bus index $n_L \in N_L$ in its corresponding bus line. Note that bus $n_L - 1$ could belong to line $l = 0$ or $l = 1$ depending on the underlying schedule.[1] If we assume i.i.d. white noise with variance σ^2 along the corridor, the variance of these holding times $D^l_{n_l,s}$ can be bounded from above by the limiting steady state case:

$$var(D^l_{n_l,s}) \leq \sigma^2_{D^l}(\boldsymbol{f}^l, \boldsymbol{f}^{R_L(n_L-1)}, \beta^l_{n_l,s}, \beta^L_{n_L,s}) \equiv \lim_{\substack{n\to\infty \\ s\to\infty}} \text{var}(D_{n,s}) \quad \text{(D.4b)}$$

$$\lim_{\substack{n\to\infty \\ s\to\infty}} \text{var}(D_{n,s}) = \sigma^2 \sum_{j=0}^{\infty}(\sum_i [(1+\beta^l_{n_l,s}+\beta^L_{n_L,s})f^l_{i|j} - \beta^l_{n_l,s}f^l_{i-1|j} - f^l_{i|j+1}]^2$$
$$+ \sum_{i'} [\beta^L_{n_L,s} f^{R_L(n_L-1)}_{i'|j}]^2)$$

and

$$d^l_{n_l,s}(\boldsymbol{f}^l, \boldsymbol{f}^{R_L(n_L-1)}, \beta^l_{n_l,s}, \beta^L_{n_L,s}) = 3\sigma_{D^l}(\boldsymbol{f}^l, \boldsymbol{f}^{R_L(n_L-1)}, \beta^l_{n_l,s}, \beta^L_{n_L,s}) \quad \text{(D.4c)}$$

One possible approach to optimize the control coefficients $\boldsymbol{f}^0$ and $\boldsymbol{f}^1$ in the two-line corridor over J control points is minimizing the weighted sum of slack times subject to some reliability constraints. The weights in the optimization correspond to the different expected unitary demand rates. The main difference with the isolated line case resides in the need to account for those passengers who can board either of the two lines. For this reason, the functions R_L and T_L previously defined play an essential role in the mathematical program that would yield the control coefficients and slacks:

[1] If the buses from lines 0 and 1 were perfectly alternated one could be sure that bus $n_L - 1$ would belong to the complement line l^c. However, this formulation is general enough to consider any other operational approach, and thus the line of bus $n_L - 1$ cannot be pre-specified.

(MP3.1)
$$\min_{\boldsymbol{f}^0,\boldsymbol{f}^1} \sum_{s=0}^{J}\sum_{l=0}^{0}\sum_{n_l=0}^{n_l^{\max}} \beta^l_{n_l,s} d^l_{n_l,s} + \sum_{s=0}^{J}\sum_{n_L=0}^{n_L^{\max}} \beta^L_{n_L,s} d^{R_L(n_L)}_{T^{-1}(n_L),s}$$
$$\text{s.t. } \sigma_\varepsilon(\boldsymbol{f}^0) \leq s^0_\varepsilon$$
$$\sigma_\varepsilon(\boldsymbol{f}^1) \leq s^1_\varepsilon$$

Where $\sigma_\varepsilon(\boldsymbol{f})$ can be obtained using expression (3.11b). Generally a transit operator would have similar reliability objective for all lines but this formulation is flexible enough to consider different reliability levels per line. There is a variety of other possible mathematical programs (e.g. minimization of average trip time in the corridor, minimization of maximum trip time, etc.) that could be used to calibrate the control coefficients and slack times for both lines. However, all formulations should consider balancing the buses' speed with their reliability. The proposed mathematical program does so since slack time is directly tied to the commercial speed in the system and the constraints impose a minimum reliability level.

Appendix E

Multi-line Demand Decomposition Methodology

To illustrate the demand decomposition methodology, let us use a simplified scenario as depicted in Figure E.1. Without loss of generality, we assume that this scenario includes two lines $l \in \{1, 2\}$ where buses travel in the same unique direction. These lines start in an overlapping corridor and split into two isolated segments. Each one of the lines has three stops, sharing the first two in the aforementioned overlapping corridor. We assume that our initial data are the hourly boarding rates for buses in each line, λ_s^{1t} and λ_s^{2t}, during a given hour period (e.g. 10am to 11am). These measurements can be expressed in vector form as $\boldsymbol{\lambda}^{1t} = [\lambda_0^{1t}, \lambda_1^{1t}, 0, 0]$. The goal is to obtain the underlying line-specific and common demand rates λ_s^1, λ_s^2, and $\lambda_s^{1,2}$.

In this system one can construct two line-specific connectivity matrices C^1 and C^2. These are sparse matrices where $C_{i,j}^1 = 1$, if stop i directly precedes stop j in line 1. In our illustration system, the connectivity matrix for lines 1 and 2 would be:

$$C^1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, C^2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

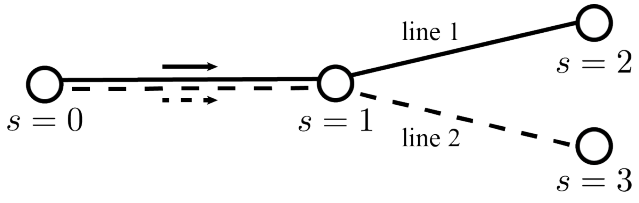


Figure E.1: Simplified two-line scenario used to illustrate the demand decomposition methodology.

These matrices can also be used to differentiate the overlapping areas in the network from the isolated ones. The connectivity matrix in the overlapping corridor can be obtained by performing the Hadamard or element wise product of C^1 and C^2. The result of performing this operation in the idealized system would be:

$$C^{1,2} = C^1 \odot C^2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

On the other hand, the connectivity matrices of the isolated portion of both lines, which we will denote as $C^{1'}$ and $C^{2'}$ can be obtained by subtracting the overlapping connectivity matrix $C^{1,2}$ from the original line specific matrices:

$$C^{1'} = C^1 - C^{1,2} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, C^{2'} = C^2 - C^{1,2} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The matrices C^1 and C^2 present which stops that lay consecutively ahead of each other along the lines. The connectivity matrix of stops that are separated by an intermediate stop along the line, which we will name "two-stop-ahead" connectivity matrix, can be obtained by squaring C^1 and C^2:

$$(C^1)^2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, (C^2)^2 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

This also applies to the common corridor connectivity matrix. In this case though, since the overlapping corridor contains only two stops the result would be a zero matrix: $(C^{1,2})^2 = 0_{4,4}$. Similarly, the "two-stop-ahead" connectivity matrix for the isolated portion of the lines would be $(C^{1'})^2 \equiv (C^1)^2 - (C^{1,2})^2 = (C^1)^2$ and $(C^{2'})^2 \equiv (C^2)^2 - (C^{1,2})^2 = (C^2)^2$. The "$n$-th-stop-ahead", for those systems that would require its calculation, could be obtained similarly. These connectivity matrices, in conjunction with the system origin/destination matrix, OD[1], and the measured demands λ^{1t} and λ^{2t} can then be used to obtain the desired

[1]Note that the O/D matrix for this system will be an upper triangular matrix with zeros in the diagonal terms.

line-specific demand parameters as:

$$(\boldsymbol{\lambda}^1)' = (\boldsymbol{\lambda}^{1t} + \boldsymbol{\lambda}^{2t})' \sum_{i=1}^{S_1} (OD \odot ((C^1)^i - (C^{1,2})^i)), \tag{E.1a}$$

$$(\boldsymbol{\lambda}^2)' = (\boldsymbol{\lambda}^{1t} + \boldsymbol{\lambda}^{2t})' \sum_{i=1}^{S_2} (OD \odot ((C^2)^i - (C^{1,2})^i)), \tag{E.1b}$$

where the sumation limit S_l refers to the maximum number of stops in each one of the lines involved in the calculation. The common demand rate would similarly be:

$$(\boldsymbol{\lambda}^{1,2})' = (\boldsymbol{\lambda}^{1t} + \boldsymbol{\lambda}^{2t})' \sum_{i=1}^{S_{1,2}} (OD \odot (C^{1,2})^i)). \tag{E.1c}$$

The values of $\beta_s^{\cdot}$ could then be computed by multiplying the average passenger boarding time t_b by the demand rates $\lambda_s^{\cdot}$. In the Dbus' system, the average boarding time per passenger was assumed to be 2 seconds.[2]

Note that these calculations assume that riders avoid unnecessary transfers, i.e. a user with an origin in the common corridor and a destination in the isolated portion of a line will only board a bus from that line. Conversely, a user could board the first bus to arrive and then transfer at some other station down the line. But given that transfers are negatively perceived [43], one should not expect to observe this behavior very frequently.

Finally we include several tables that illustrate the data inputs and outputs of this process, for a subset of the stations involved in the case study. Table E.1a provides a view of the hourly demand rates captured by Dbus' APC system in the first 9 stations of the outbound corridor for line 5 during regular workdays (Monday-Thursday) in February 2013. Table E.1b presents similar input values for the first 9 stations in the outbound direction of line 25 during that same time period.

These inputs were used to obtain the final $\beta_s^{\cdot}$ that were used in the application of the control. Tables E.2a, E.2b, and E.2c provide the resulting beta levels for that corridor. These results reveal the operational characteristics of the service in lines 5 and 25. During the morning rush, from 7am until 10am, the buses from line 5 that travel in the outbound direction have to cover the same stops as the buses in line 25 cover during all day (stops with id 239, 240, 241, 269, 277, 278, and 50). During the rest of the day, line 5 covers a different set of stops in the outbound direction and that is reflected by the common demand parameters $\beta_s^{5,25}$ for the aforementioned stops, which take 0 values during that period of time.

[2]In Dbus, 95% of their users resort to payments using a contactless card, which speeds up the boarding process significantly.

Table E.1: Dbus average hourly demand rates captured by their APC system and used in the demand calibration.

(a) Line 5, first 9 stations in the outbound direction, February 2013.

	Time of Day																	
[Id] Station Name	0-6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
[34] Boulevard 17	0	143	232	168	144	144	162	174	154	118	115	119	152	169	135	70	48	3
[35] Londres	0	56	161	99	81	94	109	128	111	87	80	83	95	126	107	46	18	4
[36] La Perla	0	5	29	18	12	13	17	31	19	15	11	27	10	17	12	5	2	1
[55] Matia	0	0	0	5	6	9	8	9	8	4	4	6	8	10	9	2	0	0
[54] Benta Berri	0	0	0	4	9	13	17	17	7	22	4	9	10	12	8	2	0	0
[53] R.M. Azkue	0	0	0	2	3	4	6	8	3	3	3	4	6	5	4	1	0	0
[52] Zarautz 60	0	0	0	1	8	1	1	1	2	0	1	0	0	2	1	0	0	0
[51] Zarautz 90	0	0	0	0	2	1	3	1	1	1	1	0	1	1	0	0	0	0
[269] Avda Zarauz 122	0	0	0	0	0	0	0	0	0	0	7	3	1	0	0	0	0	0

(b) Line 25, first 9 stations in the outbound direction, February 2013.

	Time of Day																	
[Id] Station Name	0-6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
[34] Boulevard 17	0	50	71	45	43	41	46	45	44	39	31	38	32	44	39	41	15	8
[35] Londres	0	21	37	25	27	28	29	36	31	21	23	26	28	35	23	15	17	1
[36] La Perla	0	3	6	4	4	3	6	6	4	7	5	4	3	6	3	2	1	0
[239] Zumalakarregi 10	0	3	3	1	2	3	6	2	1	3	2	3	4	3	1	0	0	0
[240] Antig Anbul	0	1	3	3	4	5	7	8	4	5	3	5	5	4	2	1	1	0
[241] Magisterio	0	2	1	5	2	2	4	3	3	2	6	3	4	3	3	1	0	0
[277] Unibertsitatea Tol.70 I	0	1	2	1	2	1	4	5	11	5	10	4	4	2	3	1	0	0
[278] Tolosa 112	0	0	1	1	1	1	1	1	2	1	2	1	1	1	2	0	1	0
[50] Tolosa 138	0	0	0	1	1	0	1	1	1	1	2	2	2	2	1	0	0	0

Table E.2: Dbus $\beta_s^{\cdot}$ demand parameters resulting from the calibration.

(a) Line 5, first 9 stations in the outbound direction, February 2013.

	Time of Day																	
[Id] Stop Name	0-6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
[34] Boulevard 17	0	6.6E-2	1.1E-1	8.1E-2	6.8E-2	6.9E-2	7.7E-2	8.4E-2	7.3E-2	5.5E-2	5.5E-2	5.6E-2	7.6E-2	8.2E-2	6.4E-2	2.8E-2	2.3E-2	0
[35] Londres	0	2.5E-2	7.9E-2	4.8E-2	3.8E-2	4.4E-2	5.3E-2	6.1E-2	5.3E-2	4.3E-2	3.8E-2	3.9E-2	4.5E-2	6.0E-2	5.3E-2	2.1E-2	5.3E-3	1.9E-3
[36] La Perla	0	1.9E-3	1.4E-2	8.9E-3	5.6E-3	6.4E-3	7.8E-3	1.6E-2	9.4E-3	6.4E-3	4.7E-3	1.4E-2	4.7E-3	7.8E-3	5.8E-3	2.2E-3	8.0E-4	6.0E-4
[55] Matia	0	0	0	2.8E-3	3.3E-3	5.0E-3	4.4E-3	5.0E-3	4.4E-3	2.2E-3	2.2E-3	3.3E-3	4.4E-3	5.6E-3	5.0E-3	1.1E-3	0	0
[54] Benta Berri	0	0	0	2.2E-3	5.0E-3	7.2E-3	9.4E-3	9.4E-3	3.9E-3	1.2E-2	2.2E-3	5.0E-3	5.6E-3	6.7E-3	4.4E-3	1.1E-3	0	0
[53] R.M. Azkue	0	0	0	1.1E-3	1.7E-3	2.2E-3	3.3E-3	4.4E-3	1.7E-3	1.7E-3	1.7E-3	2.2E-3	3.3E-3	2.8E-3	2.2E-3	6.0E-4	0	0
[52] Zarautz 60	0	0	0	6.0E-4	4.4E-3	6.0E-4	6.0E-4	6.0E-4	1.1E-3	0	6.0E-4	0	0	1.1E-3	6.0E-4	0	0	0
[51] Zarautz 90	0	0	0	0	1.1E-3	6.0E-4	1.7E-3	6.0E-4	6.0E-4	6.0E-4	6.0E-4	0	6.0E-4	6.0E-4	0	0	0	0
[269] Avda Zarauz 122	0	0	0	0	0	0	0	0	0	0	3.9E-3	1.7E-3	6.0E-4	0	0	0	0	0

(b) Line 25, first 9 stations in the outbound direction, February 2013.

	Time of Day																	
[Id] Stop Name	0-6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
[34] Boulevard 17	0	1.4E-2	2.0E-2	1.3E-2	1.2E-2	1.1E-2	1.3E-2	1.3E-2	1.2E-2	1.1E-2	8.6E-3	1.1E-2	8.9E-3	1.2E-2	1.1E-2	1.1E-2	4.2E-3	4.4E-3
[35] Londres	0	5.8E-3	1.0E-2	6.9E-3	7.5E-3	7.8E-3	8.1E-3	1.0E-2	8.6E-3	5.8E-3	6.4E-3	7.2E-3	7.8E-3	9.7E-3	6.4E-3	4.2E-3	4.7E-3	3.0E-4
[36] La Perla	0	8.0E-4	1.7E-3	1.1E-3	1.1E-3	8.0E-4	1.7E-3	1.7E-3	1.1E-3	1.9E-3	1.4E-3	1.1E-3	8.0E-4	1.7E-3	8.0E-4	6.0E-4	3.0E-4	0
[239] Zumalakarregi 10	0	8.0E-4	8.0E-4	3.0E-4	1.1E-3	1.7E-3	3.3E-3	1.1E-3	6.0E-4	1.7E-3	1.1E-3	1.7E-3	2.2E-3	1.7E-3	6.0E-4	0	0	0
[240] Antig Anbul	0	3.0E-4	8.0E-4	1.1E-3	2.2E-3	2.8E-3	3.9E-3	4.4E-3	2.2E-3	2.8E-3	1.7E-3	2.8E-3	2.8E-3	2.2E-3	1.1E-3	6.0E-4	6.0E-4	0
[241] Magisterio	0	6.0E-4	3.0E-4	2.8E-3	1.1E-3	1.1E-3	2.2E-3	1.7E-3	1.7E-3	1.1E-3	3.3E-3	1.7E-3	2.2E-3	1.7E-3	1.7E-3	6.0E-4	0	0
[277] Unibertsitatea Tol.70 I	0	0	6.0E-4	8.0E-4	6.0E-4	6.0E-4	6.0E-4	6.0E-4	1.1E-3	6.0E-4	1.1E-3	6.0E-4	6.0E-4	6.0E-4	1.1E-3	0	6.0E-4	0
[278] Tolosa 112	0	0	3.0E-4	6.0E-4	6.0E-4	0	6.0E-4	6.0E-4	6.0E-4	6.0E-4	1.1E-3	1.1E-3	1.1E-3	1.1E-3	6.0E-4	0	0	0
[50] Tolosa 138	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

(c) Common demand, lines 5 and 25, outbound direction, February 2013.

	Time of Day																	
[Id] Stop Name	0-6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
[34] Boulevard 17	0	1.4E-2	2.0E-2	1.3E-2	1.2E-2	1.1E-2	1.3E-2	1.3E-2	1.2E-2	1.1E-2	8.6E-3	1.1E-2	8.9E-3	1.2E-2	1.1E-2	1.1E-2	4.2E-3	0
[35] Londres	0	5.8E-3	1.0E-2	6.9E-3	7.5E-3	7.8E-3	8.1E-3	1.0E-2	8.6E-3	5.8E-3	6.4E-3	7.2E-3	7.8E-3	9.7E-3	6.4E-3	4.2E-3	4.7E-3	3.0E-4
[36] La Perla	0	8.0E-4	1.7E-3	1.1E-3	1.1E-3	8.0E-4	1.7E-3	1.7E-3	1.1E-3	1.9E-3	1.4E-3	1.1E-3	8.0E-4	1.7E-3	8.0E-4	6.0E-4	3.0E-4	0
[239] Zumalakarregi 10	0	8.0E-4	8.0E-4	3.0E-4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
[240] Antig Anbul	0	3.0E-4	8.0E-4	6.0E-4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
[241] Magisterio	0	6.0E-4	3.0E-4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
[277] Unibertsitatea Tol.70 I	0	0	6.0E-4	3.0E-4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
[278] Tolosa 112	0	0	3.0E-4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
[50] Tolosa 138	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0