Title
Dynamic Control of Complex Transit Systems

Authors
Argote-Cabanero, Juan
Daganzo, Carlos F
Lynn, Jacob W

Publication Date
2015-04-30

Dynamic Control of Complex Transit Systems

Juan Argote-Cabanero[a,b], Carlos F. Daganzo[a,b], Jacob W. Lynn[b]

[a] *Institute of Transportation Studies, University of California, Berkeley, CA 94720*
[b] *VIA Analytics, 2120 University Ave, Office 427, Berkeley, CA 94704*

Abstract

This paper proposes a dynamic control method to overcome bunching and improve the regularity of fixed-route transit systems. The method uses a combination of dynamic holding and en-route driver guidance to achieve its objectives. It applies to systems with a mix of headway-based and schedule-based lines but it is evaluated for scheduled systems as this is the more challenging application. Improved schedule adherence is the goal.

The method's calculation complexity per piece of advice does not increase with system size. As a result, the method is scalable and can be used with large multi-line systems, no matter how complex. When controled, each vehicle is mostly affected by exogenous disturbances (e.g. traffic) and very little by other vehicles. As a result, disruptions to a vehicle or group of vehicles caused by inattentive drivers or control equipment failures remain confined to the vehicles experiencing the problems. The control method effectively quarantines "disease".

The method is evaluated analytically and with simulations over a broad range of conditions, including schedules with zero slack. The method was also evaluated by observing the performance of a real world multi-line system that uses inexpensive on-board tablets to apply the control. The evaluation addresses driver compliance and equipment malfunction issues. It is found that the method is resilient and improves reliability considerably even under challenging conditions.

Keywords: transit reliability, real world operation, human factors, bus bunching, adaptive control, dynamic holding, multi-line control, resiliency

UCB-ITS-WP-2015-1 04-30-2015

Publications in the Working Paper series are issued for discussion and are not considered final reports.

Email addresses: juan.argote@berkeley.edu (Juan Argote-Cabanero), daganzo@berkeley.edu (Carlos F. Daganzo), jacob@v-a.io (Jacob W. Lynn)

1. Introduction

Much of the scheduling/control research on transit reliability considers isolated lines, and ignores real world complications such as intermittent communications, temporary GPS failures and inattentive drivers. These simplifications are required by the scale and complexity of typical transit operations, which are further complicated by the so called "bus (vehicle) bunching" phenomenon.[1] This paper expands current knowledge by introducing a simple control algorithm that can counterbalance bunching and improve the reliability of transit systems consisting of any number of interacting lines, no matter how many or how complex. It then evaluates the algorithm's performance recognizing the effect of the above-mentioned complications. The work builds on Argote-Cabanero (2014), which presented a proto-algorithm with some of these features and showed how to put it into practice.

Newell and Potts (1964) was the first scientific work to characterize the bunching phenomenon. It noted that a bus' travel time is not only affected by traffic and driver behavior but also by its headway with the bus it follows because long headways slow down a bus by forcing it to pick up more passengers. Building on this idea, the reference then showed that the equations of bus motion exhibit a positive feedback loop; i.e., that if a bus is delayed so that its headway increases then it is delayed even more. It further showed that this feedback loop leads to unstable platoon travel, and ultimately to bunching.

To compensate for this bunching instability, and to allow lagging buses to get back on schedule, bus operating agencies often include slack into their timetables. Drivers are asked to compare the current time against the schedule at designated control points, and to hold if they would otherwise depart early. An early analysis of this idea can be found in Osuna and Newell (1972). Unfortunately, schedule-based control methods usually require too much slack to be effective. For this reason, the slack time provided in real systems is usually smaller, or nil, and reliability suffers.

Because schedule-based methods are myopic – each bus is controlled without considering what others are doing – more systemic approaches have been

[1] The word "bus" in this paper shall stand for any transit vehicle; e.g., a tram or train.

explored. Some works (Eberlein et al., 2001; Sun and Hickman, 2005; Delgado et al., 2009; Liu et al., 2013) have proposed dynamic control strategies based on short term predictions of the system's evolution. Holding times are optimized with rolling horizon methods that consider real-time data for all the buses simultaneously. Although this is a considerable improvement over schedule-based control, the new methods' calculation complexity increases rapidly with system size. This restricts the domain of application to relatively small systems. Complexity also limits the frequency with which drivers can receive advice. So it is important that the advice is good for the time between refreshes. Unfortunately, this could be a challenge if traffic and drivers do not behave as anticipated in the rolling horizon calculations.

A second type of approach that alleviates some of these drawbacks replaces predictions with adaptation to the present. Works in this genre include Daganzo (2009), Daganzo and Pilachowski (2011) and Xuan et al. (2011), which used control theory to develop strategies for a single line that could be operated either with a schedule or based on headways; and Bartholdi and Eisenstein (2012), which used Markov chains to develop a rule for single lines operated on headways. All the works in this adaptive genre are characterized by using neither predictions nor detailed optimizations. Instead, they use simple preventive principles that can be refreshed frequently to prevent bunching before it occurs. As such, they promise to be robust to errors in the assumptions, and to be scalable.

To our knowledge, no adaptive works exist at present which have exploited this scalability to consider systems of multiple interacting lines. Therefore, the present paper attempts to fill this void. It develops and evaluates a simple control rule that can be applied to any subset of a set of interacting lines – no matter how large or complex. Control is exercised through dynamic holding at stations, and by providing drivers with on board guidance while cruising. Cruising guidance is useful because it enables drivers to absorb holding time while en-route, reducing the need for holding at stations and (as will be shown) increasing efficiency. Cheap mobile technology makes this control feature economical and feasible.

Theoretical predictions of the control's performance are compared both with simulations and with the actual performance of a real bus system. Other empirical evaluations of different control methods can be found in Bartholdi and Eisenstein (2012) and Lizana et al. (2014). These two works, however, deal with isolated lines and neither incorporates cruising guidance. The former tests an adaptive algorithm for headway-based lines and the latter a

predictive algorithm in the rolling horizon family. These experiments lasted 1 and 2 days, respectively. The tests in this paper last for months, include cruising guidance, and involve two interacting lines with an underlying schedule – both controlled. As such the tests extend and enrich the previous empirical findings. The real system of the test also included some non-cooperative drivers and temporary but recurrent failures in IT services; thus the tests also speak to the proposed approach's robustness.

For maximum generality, the paper focuses on systems run with a schedule.[2] It is organized as follows. To begin, Sec. 2 introduces background material, including an existing algorithm for isolated lines; and Sec. 3 unveils several properties of this algorithm that facilitate its generalization to systems of interacting lines. Then, Sec. 4 presents this generalization; Sec. 5 examines the generalized algorithms' performance under a broad range of conditions, including schedules with zero and even negative slack; and Sec. 6 examines its resiliency to disruptions. Finally, Section 7 provides performance predictions for applications where, in addition to holding advice at stations, drivers also receive cruising guidance while en route; Sec. 8 presents the results of several simulations and field tests of a real system; and Sec. 9 presents some concluding thoughts.

2. Background

This section outlines the modeling framework for control of isolated lines, including notation and some known results. Section 2.1 describes the problem and formulates its basic equations. The theory is developed assuming that the slack in the schedule is ample enough to absorb the (small) schedule deviations that arise despite the control. This keeps the models "linear" so that analysis is possible. Section 2.2 then describes a simple form of control that will play an important role in this paper.

2.1. Problem definition

The assumptions and notation are as in Daganzo (2009). Considered is a single line with a given schedule. Its stations are denoted s and numbered

[2]The results in this paper can be also applied to systems run on headways merely by introducing an underlying virtual schedule. The reverse is not true, however. Control methods designed to guarantee even headways do not necessarily guarantee schedule adherence.

consecutively, $s = 0, 1, \cdots$, in the direction of travel. Likewise, bus trips along the line are indexed by n, which increases as $n = 0, 1, \cdots$ with the time of day. With this in mind, the following data are given:

- $t_{n,s}$: scheduled arrival time of bus n at station s.
- $H_{n,s}$: the target headway. ($H_{n,s} = t_{n,s} - t_{n-1,s}$, $\forall n > 0$.)
- $c_{n,s}$: the (cruising) time bus n is expected to take from its departure from station s to its arrival at $s + 1$, considering the time of day.
- $\beta_{n,s}$: Expected increase in the dwell time for bus n at station s due to additional passenger boardings when its headway increases by one time unit. (As explained in Daganzo (2009), $\beta_{n,s}$ is a dimensionless constant that typically ranges from 10^{-2} to 10^{-1}.)
- $d_{n,s}$: the amount of pre-planned slack included in the schedule of bus n at station s.

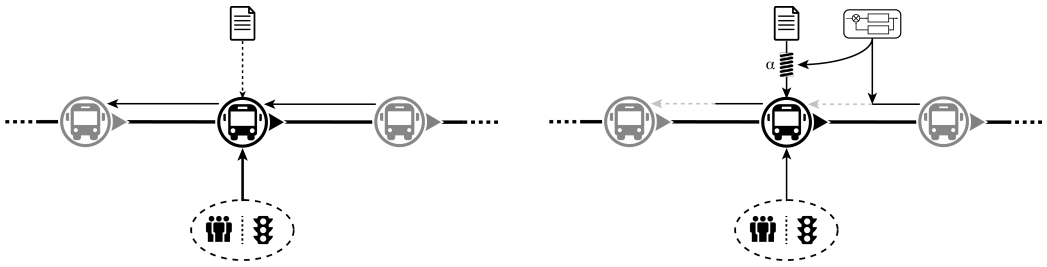
The above defines the schedule and the buses' average behavior. Randomness is introduced as follows:

- $\nu_{n,s+1}$: zero-mean, independent, exogenous random noise in the trip time of bus n from s to $s + 1$.
- $\sigma^2_{n,s+1}$: variance of $\nu_{n,s+1}$. (Given parameter.)

The exogenous noise can be due to traffic, driver error and to variability in passenger demand. It affects travel unpredictably and prevents buses from being on schedule.

Figure 1a presents an influence diagram indicating how not just this noise, but also the bus in front affects an uncontrolled bus' travel, as per the headway effect in Newell and Potts (1964). The dark horizontal arrows denote the direction of movement and the thin solid arrows causality. Note how thin arrows pointing to the bus emanate both from the bus in front and from the elliptical bubble at the bottom which represents the exogenous noise. The dotted arrow between the schedule at the top and the bus stresses that the schedule is not used if the bus runs uncontrolled.

The job of a control algorithm is compensating for both of the influences in Fig. 1a so as to keep the schedule deviations as small as possible. A control



(a) Uncontrolled line. (b) Controlled line.

Figure 1: Causal relations affecting a bus' motion in an isolated line: (a) no control; (b) with control.

mechanism can be qualitatively depicted by adding arrows. For example, an arrow pointing from the schedule to the bus (replacing the dotted line) could be used to represent schedule-control, since this control method only uses the schedule to affect the bus. Part (b) of the figure represents the form of adaptive control discussed later in this paper.

The following state and control variables are used to model the buses' motion:

- $a_{n,s}$: actual arrival time of bus n at station s. (State variable.)
- $\varepsilon_{n,s} \equiv a_{n,s} - t_{n,s}$: schedule deviation of bus n at station s. (State variable: positive values are lateness and negative ones earliness.)
- $h_{n,s} \equiv a_{n,s} - a_{n-1,s}$: actual headway between bus n and its leading bus, $n-1$, at station s. (State variable.)
- $D_{n,s}$: actual holding time of bus n at station s. (Control/decision variable.)

Note from the above definitions that the line's arrival-based schedule satisfies:

$$t_{n,s+1} = t_{n,s} + \beta_{n,s} H_{n,s} + d_{n,s} + c_{n,s}. \tag{1}$$

The actual arrival times satisfy an expression similar to (1), which uses the actual headways and holding times, and includes the noise:

$$a_{n,s+1} = a_{n,s} + \beta_{n,s} h_{n,s} + D_{n,s} + c_{n,s} + \nu_{n,s+1}. \tag{2}$$

This expression implies that an uncontrolled bus' travel time between stations is of the form, $a_{n,s+1} - a_{n,s} = \beta_{n,s} h_{n,s} + c_{n,s} + \nu_{n,s+1}$. The first term, $\beta_{n,s} h_{n,s}$, captures the endogenous influence of the previous bus depicted by the left-pointing arrow of Fig. 1a, and the other two terms, $c_{n,s} + \nu_{n,s+1}$, the exogenous influences depicted by the vertical arrow.

The constants $H_{n,s}$ and $c_{n,s}$ appearing in the previous two equations can be eliminated from the formulation by subtracting (1) from (2), and then expressing headways in terms of schedule deviations. With our definitions of $\varepsilon_{n,s}, H_{n,s}$ and $h_{n,s}$, these steps yield:

$$\begin{aligned} \varepsilon_{n,s+1} &= \varepsilon_{n,s} + \beta_{n,s}(h_{n,s} - H_{n,s}) + (D_{n,s} - d_{n,s}) + \nu_{n,s+1} \\ &= \varepsilon_{n,s} + \beta_{n,s}(\varepsilon_{n,s} - \varepsilon_{n-1,s}) + (D_{n,s} - d_{n,s}) + \nu_{n,s+1}. \end{aligned} \tag{3}$$

The system dynamics are fully specified when $D_{n,s}$ in the above expression is replaced by a non-negative function of all past system-wide schedule deviations. Such a function is the holding control algorithm. The following subsection describes a particular algorithm (Xuan et al., 2011) that will be generalized in this paper to systems of interacting lines. The subsection unveils a unique property of this algorithm that makes the generalization practical.

2.2. A simple control

Considered here is the so-called "simple control" in Xuan et al. (2011). Its holding time expression has a single degree of freedom – a parameter, α, which is required to be in the $(0, 1)$ interval to ensure stability. The expression is:

$$D_{n,s} = \max\{0, \beta_{n,s}\varepsilon_{n-1,s} + (\alpha - 1 - \beta_{n,s})\varepsilon_{n,s} + d_{n,s}\} \quad \text{where} \quad \alpha \in (0, 1). \tag{4}$$

The model is called "simple" because if the slack values $d_{n,s}$ are large enough so the second argument of the maximum operation in (4) stays positive, then: (i) the holding time expression is linear,

$$D_{n,s} = \beta_{n,s}\varepsilon_{n-1,s} + (\alpha - 1 - \beta_{n,s})\varepsilon_{n,s} + d_{n,s} \quad \text{where} \quad \alpha \in (0, 1); \tag{5}$$

and (ii) the dynamic equation simplifies to:

$$\varepsilon_{n,s+1} = \alpha\varepsilon_{n,s} + \nu_{n,s+1} \qquad \text{where} \quad \alpha \in (0, 1). \tag{6}$$

3. Preliminary Ideas

This section unveils some useful features of the simple control. Of particular interest are the decoupled bus dynamics that can be seen from (6) in the linear regime. This will turn out to be useful, and therefore is the first feature explored.

To isolate buses from one another while stabilizing their trajectories, the control term $D_{n,s}$ must cancel out the (endogenous) influence from other buses, and then smooth out the exogenous noise. Figure 1b illustrates this idea by means of a controller box with two emerging arrows that correspond to these two functionalities: a vertical arrow that isolates the bus and a curved arrow that stabilizes it.

Mathematically, isolation in the linear regime is achieved by including in the holding time a term, denoted $D^{(i)}_{n,s}$, equal to the negative of all the terms in (3) that involve other buses; i.e., $D^{(i)}_{n,s} = \beta_{n,s}\varepsilon_{n-1,s}$. Stabilization is then achieved with a term, $D^{(ii)}_{n,s}$, that cancels out the bus' own endogenous terms in (3) and replaces them by $\alpha\varepsilon_{n,s}$. Consideration of (3) shows that $D^{(ii)}_{n,s} = (\alpha - 1 - \beta_{n,s})\varepsilon_{n,s} + d_{n,s}$.

Since the only dynamic variable in this last expression is $\varepsilon_{n,s}$, i.e., the difference between the actual and scheduled arrival times of bus n, the curved arrow of Figure 1b points to the connection between the schedule and the bus. A spring with the label "α" is used as an icon to represent the result of this connection, as encapsulated in (6).

Explored next, is the relationship between α and the magnitude of the schedule deviations in the linear regime. The metric used to quantify these magnitudes is the expectation of the square of $\varepsilon_{n,s}$, denoted $\langle\varepsilon_{n,s}\rangle$.

To do this, assume that all the buses are initially dispatched on time, i.e., $\varepsilon_{n,0} = 0\ \forall n$. Then, recursively replace $\varepsilon_{n,s}$ on the RHS of (6) by its expression in terms of $\varepsilon_{n,s-1}$, again using (6), and then iterate this recursion until $s = 0$. The result is:

$$\varepsilon_{n,s+1} = \alpha^{s+1}\varepsilon_{n,0} + \sum_{i=0}^{s}\alpha^i\nu_{n,s+1-i} = \sum_{i=0}^{s}\alpha^i\nu_{n,s+1-i} \quad \text{where} \quad \alpha \in [0,1). \tag{7}$$

Since the terms of the second sum in (7) are independent random variables with zero mean, we can write:

$$\langle\varepsilon_{n,s+1}\rangle = \mathrm{var}(\varepsilon_{n,s+1}) = \sum_{i=0}^{s}\alpha^{2i}\sigma^2_{n,s+1-i}. \tag{8}$$

This is the desired relationship.

Note that the RHS of (8) is a uniformly bounded quantity if the $\sigma_{n,s}$ are uniformly bounded. Thus, the buses operate in a stable manner (i.e. with bounded deviations) no matter how long the route or how many buses operate on it. In the homogeneous case where $\sigma_{n,s}^2 \equiv \sigma^2$ the summation is a geometric series, and therefore the bound is:

$$\langle \varepsilon_{n,s} \rangle = \mathrm{var}(\varepsilon_{n,s}) \leq \frac{\sigma^2}{1-\alpha^2} \quad \forall n, s, \quad \text{where} \quad \alpha \in (0,1). \tag{9}$$

It is evident from (8) and (9) that if the system is in the linear regime, the smaller α the better the control. Thus, if the schedule has so much slack that the system is in the linear regime for $\alpha = 0$, then $\alpha^* = 0$.[3] However, in most real cases the schedules are not that lax. Thus, reliability is maximized in these cases by an $\alpha^* > 0$ that must trigger some non-linearities; and the tighter the schedule the more non-linear the behavior and the higher α^*.

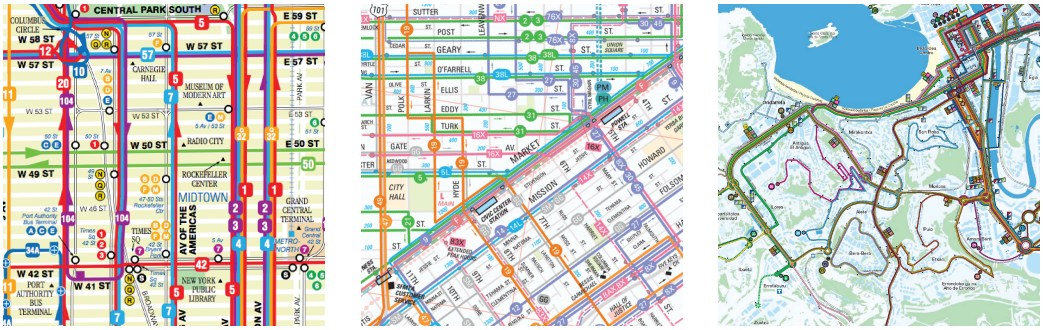
4. Systems of Interacting Lines

In the real world the situation is often more complex than the isolated line discussed in Secs. 2 and 3. Fig. 2 illustrates the common case of interlining, where buses are not just affected by their intra-line headways but also by their headways with buses from other lines. This inter-line coupling effect is triggered by passengers who are indifferent between multiple lines that share a station, and by transfers. In view of this, the simple control is now generalized to systems of interacting lines. It will be shown that all the results from Secs. 2.2 and 3 apply to the generalization.

To capture multi-line interactions generally and directly, the connection between a bus' expected dwell time and the number of boarding passengers is modeled as an arbitrary function of the schedule deviations of the bus in question and those of all previous buses from all lines to have visited the station.[4] In the case of an isolated line this function was specified as

[3] Consideration of (3 -6) shows that the selection $\alpha = 0$ corresponds to the conventional form of schedule control. In this context, the linear regime arises if there is so much slack that buses always arrive early at every station so they are always held for a positive amount of time. To our knowledge, no bus system operates in this way.

[4] Passenger counts and origin/destination tables are not used because these data are less reliable and can only be integrated into a model of dwell time with additional assumptions that could introduce systematic errors.



(a) MTA, New York City, USA.

(b) Muni, San Francisco, USA.

(c) Dbus, San Sebastián, Spain.

Figure 2: Real world bus systems with multi-line corridors.

$\beta_{n,s}(\varepsilon_{n,s} - \varepsilon_{n-1,s})$. In the general case it will be expressed as: $\beta_{n,s}\varepsilon_{n,s} - F_{n,s}(B_{n,s}, E_{n,s})$, where $\beta_{n,s}$ has the same interpretation as for a single line, $F_{n,s}$ is an arbitrary function that should make physical sense, $B_{n,s}$ is a set of parameters and $E_{n,s}$ is the set of deviations by all buses that have visited s prior to n.[5]

With these conventions, the dynamic equation for bus n of a non-isolated line becomes:

$$\varepsilon_{n,s+1} = \varepsilon_{n,s} + \beta_{n,s}\varepsilon_{n,s} - F_{n,s}(B_{n,s}, E_{n,s}) + (D_{n,s} - d_{n,s}) + \nu_{n,s+1}. \qquad (10)$$

The term containing $F_{n,s}$ captures the influence of all other buses on bus n.

Recall from Fig. 1b that the holding control expression for isolated lines was the sum of an isolation term that canceled the influence of all other buses on each bus, and a stabilization term that involved only data from the bus in question. This structure allows buses from an isolated line to operate independently from one another in the linear regime.

In the present case, the same effect can be achieved by modifying the isolation term so it now cancels the endogenous influences from all relevant buses from any line, as illustrated by Fig. 3; and then leaving the stabilization term unchanged. As the figure suggests, such extended framework must exhibit identical linear dynamics as the original, including the bus independence property.

[5]Appendix Appendix A shows how these parameters can be estimated from the buses run-time data harvested during operations.

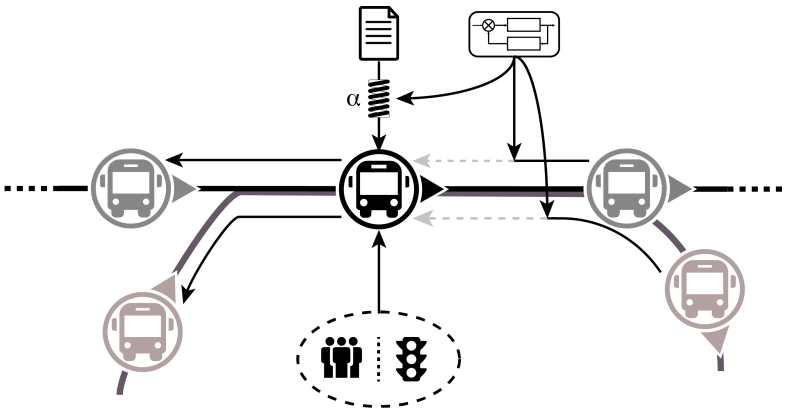


Figure 3: Causal relations affecting the buses' motion in a non-isolated line.

This can be verified algebraically. Since $D_{n,s}^{(i)}$ now must cancel the influence of all buses on bus n (not just the influence of the previous bus from its own line to have visited the station) it has the form: $D_{n,s}^{(i)} = F_{n,s}(B_{n,s}, E_{n,s})$. And since the stabilization term $D_{n,s}^{(ii)}$ is left unchanged (i.e., $D_{n,s}^{(ii)} = (\alpha - 1 - \beta_{n,s})\varepsilon_{n,s} + d_{n,s}$) the complete holding time becomes:

$$D_{n,s} = F_{n,s}(B_{n,s}, E_{n,s}) + (\alpha - 1 - \beta_{n,s})\varepsilon_{n,s} + d_{n,s}. \tag{11}$$

Now, obtain the dynamic equation by combining (11) and (10). The result is still (6), as claimed. Thus, everything (6) implies, including (7 - 9), continues to apply for the generalized algorithm.

Note that the derivation assumed nothing about the nature of the operation of the other lines; each can be controlled or uncontrolled. Lines can even suffer from bunching. The only requirement is that the function $F_{n,s}$ and its arguments $B_{n,s}$ be known; and that the arguments of $E_{n,s}$ be recorded and made available to the controller in real time. The technology for this already exists. The control method's ability to easily deal with diverse, complex scenarios stems from its non-forecasting nature which, unlike rolling horizon methods, obviates the need for producing in real time complex, multi-line predictions. The generalized version of the simple control only requires observation of what is.

5. Reliability

The system's reliability with the proposed control method is now evaluated. The evaluation extends beyond the linear regime that has been our

focus so far. As such, the evaluation considers both, tight and lax schedules, including situations where there is so little slack and the headways are so short that it is impossible to find an α that will prevent, not just the target holding time (5) from going negative, but also buses from catching up. It shall be assumed that buses that catch up do not leapfrog.

Catch-ups can be incorporated into (2) and (3) using the FIFO rule; i.e., by requiring that $a_{n,s+1} \geq a_{n-1,s+1}$, or equivalently $\varepsilon_{n,s+1} \geq \varepsilon_{n-1,s+1} - H_{n,s+1}$. With this in mind, (3) is rewritten as the maximum of its RHS and $\varepsilon_{n-1,s+1} - H_{n,s+1}$. Therefore, the system's dynamic equations for our evaluation are:

$$\varepsilon_{n,s+1} = \max\{\varepsilon_{n-1,s+1} - H_{n,s+1}, \\ \varepsilon_{n,s} + \beta_{n,s}(\varepsilon_{n,s} - \varepsilon_{n-1,s}) + (D_{n,s} - d_{n,s}) + \nu_{n,s+1}\}, \quad (12)$$

$$D_{n,s} = \max\{0, \beta_{n,s}\varepsilon_{n-1,s} + (\alpha - 1 - \beta_{n,s})\varepsilon_{n,s} + d_{n,s}\} \quad \text{where} \quad \alpha \in (0,1). \quad (13)$$

Because (12) and (13) are cumbersome to analyze, they were simulated instead. For added realism the simulations used not just buses as agents, but also passengers. This agent-based framework, which is fully described in Argote-Cabanero (2014), allowed us to model passenger arrivals explicitly and accurately as Poison processes, instead of implicitly and approximately as in (12) and (13).[6]

The results described below are restricted to the homogeneous case. This enabled us to explore systematically a broad array of scenarios by varying just a few parameters. Note from (12) and (13) that the parameters H, β, σ^2 and d suffice to define a homogeneous system. We also choose to measure time in units of σ so that the parameter $\sigma \equiv 1$ is eliminated. Thus, the space of homogeneous systems is comprehensively described by H, β and d.

Each simulation run, representing 1 day of continuous operation, tracks 100 punctually-dispatched buses over 30 stations and records the bus arrival deviations at $s = 29$, the line's terminus. The root mean square, z, of these 100 deviations, $z \equiv \langle \varepsilon_{n,29} \rangle^{1/2}$, is used as the daily evaluation metric. To improve statistical significance this daily metric is then further averaged over 30 runs, to get a monthly average, $\bar{z}$.

For each scenario (H, β and d), the simulations were run under three control schemes: (i) uncontrolled; (ii) control by schedule; and (iii) simple control. In (i), it is assumed that drivers do not hold at any stations until

[6]This is a minor point because the approximation is very good; see Daganzo (2009).

arriving at the termini; this is modeled by using $D_{n,s} = 0$ instead of (13) in (12). In (ii), drivers who are ahead of schedule hold at two intermediate control points ($s = 9$ and $s = 19$) until they are back on schedule; this is modeled by using $D_{n,s} = 0$ for all stations that are not control points, and using (13) with $\alpha = 0$ for the control points. In (iii) drivers hold at every station, and $D_{n,s}$ is given by (13) at every station.

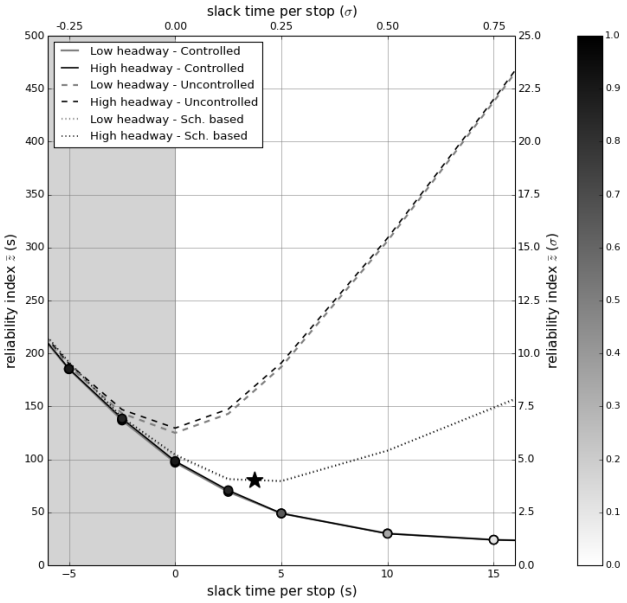
Schemes (i) and (ii) are benchmarks representing traditional operations against which the proposed control method (iii) is to be compared. Although scheme (i) is a worst-case representation of traditional operations, the scheme is sometimes used for systems with little or no slack, or systems with a small number of buses. With its closely spaced control points, scheme (ii) is a best-case representation of traditional operations.[7]

Twenty eight scenarios are considered allowing for: (i) 2 values of H (15, 30) signifying high and low frequency; (ii) 2 values of β (0.01, 0.05) signifying low and high demand; and (iii) 7 values of d (-0.25, -0.125, 0, 0.125, 0.25, 0.50, 0.75) capturing a wide range of schedules, from overly tight to somewhat relaxed. Non-linear effects turn out to play a significant role in all the scenarios. Recall that H and d are measured in units of σ. Thus, for the special case where $\sigma = 20s$ which is typical of many applications, $H = (300, 600)s$ and $d = (-5, -2.5, 0, 2.5, 5, 10, 15)s$.

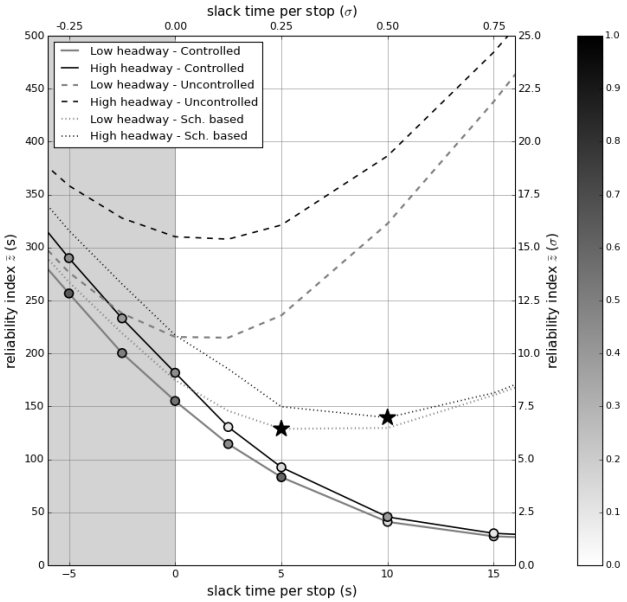
Figure 4a depicts the set of scenarios with low demand. It charts $\bar{z}$ vs d for the three control schemes, and the high and low headway scenarios. The scales running along the top and right sides of the chart express time in multiples of σ. The bottom and left scales express it in seconds for the special case where $\sigma = 20s$. Dashed lines are used for scheme (i), dotted lines for scheme (ii) and solid lines for (iii). The circles on the solid lines represent the simulation results for the simple control. The shading inside each circle corresponds to the value of α^* which, as was anticipated, declines as the slack increases. The highest reliability with schedule control is achieved at the minima of the curves for scenario (ii), marked by little stars.

Note how in Fig. 4a the curves for the high and low headway scenarios are nearly superimposed, i.e., the results for low demand are not influenced by the headway. Since the headway parameter only enters the simulation model through the no-passing rule (12), this means that said rule seldom

[7]Typical systems space their control points more widely because more control points require more field inspectors and increased work for dispatchers.



(a) Low demand ($\beta = 0.01$)



(b) High demand ($\beta = 0.05$)

Figure 4: Reliability index $\bar{z}$ obtained under three different control schemes for a variety of slack times d per station, demand levels β, and headways H.

comes into play; i.e., that bus catch-ups are infrequent when demand is low.[8]

More importantly, note from Fig. 4a how the simple control outperforms the two benchmarks. In particular note from the two little stars, which are superimposed in this figure, how it improves on the best that schedule control can achieve by about 25% .

Figure 4b depicts the scenarios with high demand using the same conventions. In the present case the non-linear effects due to bus catch-ups play a significant role. This is revealed by the high and low headway curves, which are now separated. Note from these separations how the catch-up frequency declines with increasing slack, and how it is reduced by the proposed control. Note as well that the high-headway curves lie above the low-headway curves. This occurs because catch-ups, which are more frequent with lower headways, have the beneficial effect of slowing down the buses that do the catching, which are usually ahead of schedule. This braking effect curbs the growth of earliness and reduces $\bar{z}$.

As occurred in the low demand case, the simple control continues to outperform the two benchmarks. And the improvement is more significant now as can be seen from Fig. 4b. The simple control improves on the best that schedule control can offer by about 38% for low headways and by about 68% for high headways. In fact, the simple control now achieves a significant improvement even when there is no slack.

The results in Fig. 4 give the optimum α for any given slack under a variety of conditions. As such, the figure can be used by a transit provider to optimize the control when the schedule is given. The charts can also be used as a planning tool to optimize the schedule/slack, assuming that the system is optimally controlled. This requires some care because a reduction in slack reduces travel time, which is a good outcome, but as the figure reveals it also degrades reliability. These conflicting outcomes affect in well-known ways the passenger in-vehicle travel times, waiting times, and operator costs, which should be the ultimate basis for the final choice. Since this is standard textbook material, this issue is not explored further.[9]

[8]The other non-linear effect–negative target holding times– arose frequently.

[9]Details can be found in some of the early works on single lines, which take the planning approach and jointly optimize α and d (Daganzo, 2009; Daganzo and Pilachowski, 2011; Xuan et al., 2011). These works are limited, however, because they only consider the linear regime.

6. Resiliency to Disruptions

The bus isolation property of the proposed control method suggests that disruptions to a group of buses operating in the linear regime should remain confined to the buses experiencing them. Since this is a good thing, this section explores the issue. It analyzes the linear behavior of both temporary and sustained disruptions in Sec. 6.1, and then shows in Sec. 6.2 that these analytic predictions are approximately reproduced under realistic conditions that involve the non-linear regime.

6.1. Abundant slack: linear theory

Considered here are disturbances (small disruptions) to systems with so much slack that they operate in the linear regime during and after the disturbance.

Analyzed first is the dissipation mechanism of a temporary disturbance to a single bus. To this end let m be the bus that experiences the disturbance, and assume without loss of generality that m is at station $s = 0$ when the disturbance ends. Also, let $e_{m,0}$ be the magnitude of the disturbance to the bus' schedule deviation at $s = 0$; i.e., so that if $\varepsilon_{m,0}$ is the deviation that would have arisen without the disturbance, then the actual deviation is $\varepsilon_{m,0} + e_{m,0}$.

Now, consider the difference between the middle member of (7) evaluated with two sets of initial deviations at $s = 0$ – one with and one without the disturbance to bus m. This difference is the magnitude of the residual disturbance downstream. As the reader can verify, it is 0 for all $n \neq m$. This confirms that the disturbance effects remain confined to bus m. For $n = m$ it can be seen that the difference between the two expressions is $\alpha^{s+1} e_{m,0}$, and that the difference between the expressions' second moments is $\alpha^{2(s+1)} e_{m,0}^2$. The latter means that the increase in the schedule deviation's mean square error caused by the disturbance declines by a factor α^2 with every succeeding station. For typical values of α, these residual effects are largely dissipated in less than a dozen stops.

Analyzed now is the systems' (linear) behavior under lasting disturbances to a single bus, m, e.g., as could arise with inattentive drivers. A lasting disturbance is defined by a time series of disturbances $\{e_{m,s}\}$ for $s \geq 0$. The disturbances do not have to have zero mean. This allows us to model drivers that are consistently fast or slow.

The disturbances are modeled by subsuming them in the noise term of (6) so that the dynamic equation for bus m continues to be (6), with its noise term replaced by $\nu_{m,s+1} + e_{m,s+1}$. The noise terms of all other buses remain the same. Now, since the simple control ensures that a bus' schedule deviations only depend on its previous deviations, as per (6), it follows that the effects of lasting disturbances, just like those of the temporary kind, remain confined to bus m. They do not affect other buses.

The disturbance's effect on bus m can be evaluated by introducing the new noise terms in the RHS of (7) and repeating the logic leading to (9). The result is particularly simple if the disturbances are uncorrelated with one another and with the ν's. In this case the expectation of a squared sum is the sum of the expected squared terms, so we have: $\langle \varepsilon_{n,s+1} \rangle = \sum_{i=0}^{s} \alpha^{2i} (\sigma^2_{n,s+1-i} + \langle e_{n,s+1-i} \rangle)$. This quantity is uniformly bounded if the variances and expectations on the RHS are bounded. In the uniform case, where no subscripts are needed the bound is simple:

$$\langle \varepsilon_{n,s} \rangle \;\leq\; \frac{\sigma^2 + \langle e \rangle}{1 - \alpha^2} \qquad \forall n, s, \quad \text{where} \;\; \alpha \in (0, 1). \tag{14}$$

The analysis just concluded for single-bus disturbances also sheds light on multi-bus disturbances. It implies that group disturbances must remain confined to the affected buses, and that they must decay individually as if the others did not exist. This happens because as we have just seen and was qualitatively depicted in Fig. 1b, the simple control actively isolates buses, which prevents the spread of disease. Consideration of general expressions (3 - 6) shows that this disease-containment feature is a defining property of the simple control. No other form of control achieves this feat.

6.2. Insufficient slack and non-linear effects

If the system does not have sufficient slack to stay in the linear regime before, during and after the disruptions, the proposed control may not perfectly cancel all cross-bus interactions. Single-bus disruptions could then leak to neighboring buses.

To explore this issue let us examine with the simulation tool of Sec. 5 what happens to a platoon of 5 buses, $n = \{0, 1, 2, 3, 4\}$, as it travels over a 30-station homogeneous line when the middle bus, $m \equiv 2$, loses GPS reception for a stretch of 10 stations. A loss of GPS can be more serious

than a communications failure because it also affects the following bus.[10]

One scenario from each part of Fig. 4 was considered, representing low and high demand levels: $\beta = \{0.01, 0.05\}$. The remaining parameters are common to both scenarios: $H = 300$s; $\sigma = 20$s; and $d = 5s$. They characterize a system with frequent service, moderate noise and low slack, which together imply considerable non-linearity. Used for each scenario were the corresponding near-optimum values α, which are reflected in the figure, $\alpha^* = \{0.8, 0.6\}$.

Figure 5 depicts the results of these simulations. It charts the average of $\langle \varepsilon_{n,s} \rangle^{1/2}$ across 100 different runs vs s for the the five different n. The shaded portion is the time window when bus 2 lacks GPS. For this reason, the line for this bus diverges from the rest.

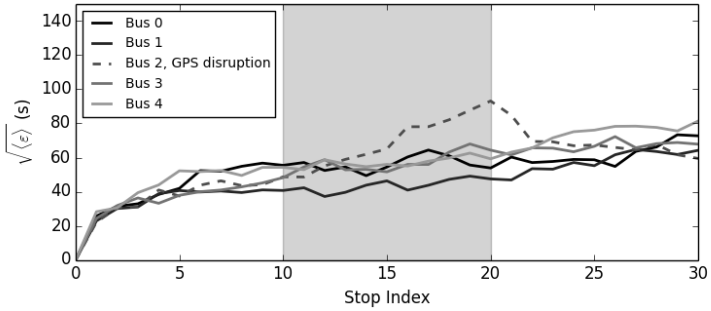
Note the little effect the disruption has on all other buses (including the following bus $n = 3$), and how it dissipates quickly after it is over. The disruption is reduced considerably in 3 stations when $\alpha = 0.8$ and in 2 stations when $\alpha = 0.6$, in agreement with the predicted decay rates for these α's, which are 0.64 and 0.36 per station.

The results do not show any effect on bus 3 even though some would be expected. For this effect to be noticeable, the deviation of bus 2 would have to have grown to be considerably larger than observed, which did not happen because the GPS failure was too short. Slight disturbances to the following bus can be seen in simulations of permanent GPS failures (Argote-Cabanero, 2014). In all cases simulated, however, the effects remain confined to at most two buses.

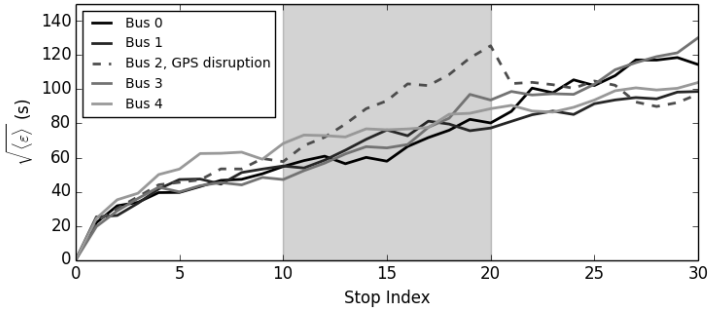
7. Cruising Guidance

So far in this paper it has been assumed that control is achieved by holding buses at stations. But on-board devices used to provide drivers with this information can also be used to give them additional information while cruising between stations. Figure 6 shows a possible way of doing this using colored bars that move up and down. It seems intuitive that if drivers are able

[10] Without GPS, the time-space coordinates of bus m are unknown so that its schedule deviation, $\varepsilon_{m,s}$, cannot be updated. As (5) reveals, this unknown quantity is an input to the holding times of both, bus m and bus $m+1$, with a stronger influence on the former than the latter. Thus, GPS failures should be expected to affect buses in pairs, and mainly the leading bus. Failures can and will be handled in simulations by calculating holding times using the most recent observed value for any missing deviation.



(a) Low demand $\beta = 0.01$.



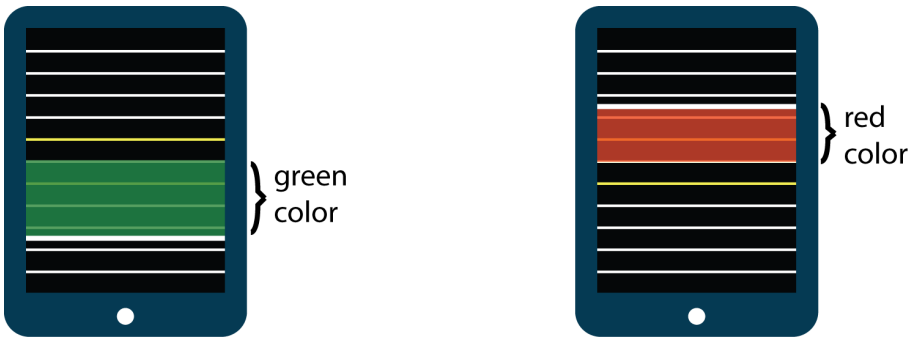
(b) High demand $\beta = 0.05$.

Figure 5: Average simulated $\sqrt{\langle \varepsilon \rangle}$ across 100 simulation runs of a GPS disruption in bus $n = 2$ of a 5-bus platoon. The GPS disruption duration is highlighted by a gray shading.

to vary their average speed just a little based on this type of cruising guidance then schedule adherence may improve. This section presents approximate formulae that quantify this effect for non-isolated lines regulated with the simple control and operating in the linear regime.

It is assumed as a first approximation that drivers are identical and that their stimulus/response behavior while traveling between stations is consistent. The stimulus is quantified numerically by a numerical score ranging from $+5$ (when the display is intense red signifying to slow down) to -5 (for intense green signifying to speed up safely). The drivers' responses are quantified by the number of extra seconds, Δ, cruising between stations. This is a negative number when the drivers speed up.

When designing the cruise control, one is free to choose any relationship



(a) Behind schedule ($\varepsilon_n > 0$). (b) Ahead schedule ($\varepsilon_n < 0$).

Figure 6: Cruising guidance on board device interface.

whatsoever between the system's state variables and the stimulus presented to the driver. In our tests, the stimulus used for the bus trip from s to $s+1$ was assumed to be $-\varepsilon_{n,s}$ (expressed in minutes). The resulting stimulus/response relation can be measured experimentally and plotted as a curve that passes through the origin. While this curve is expected to be slightly S-shaped, it will be taken to be linear in the analysis below. Thus, the added cruising time for the bus trip from s to $s+1$ is approximated by $\Delta_{s,n} = -\tau_s \varepsilon_{n,s}$, for some $\tau_s \geq 0$.

To model cruising guidance, it is assumed that the control (11) is applied to the following generalization of (10), which includes an extra term for the cruising delay, $\Delta_{s,n} = -\tau_s \varepsilon_{n,s}$:

$$\varepsilon_{n,s+1} = (1 + \beta_{n,s} - \tau_s)\varepsilon_{n,s} - F_{n,s}(B_{n,s}, E_{n,s}) + (D_{n,s} - d_{n,s}) + \nu_{n,s+1}. \quad (15)$$

Then, on combining (11) and (15) we find the following cruising-guidance version of (6):

$$\varepsilon_{n,s+1} = (\alpha - \tau)\varepsilon_{n,s} + \nu_{n,s+1} \qquad \text{where} \quad \alpha - \tau \in (0, 1). \quad (16)$$

This expression has the same structure as (6), with the smaller coefficient $(\alpha - \tau)$ playing the role of α. The analogous structure means that the control with cruising guidance continues to confine small disturbances to the bus that generates them. The reduced coefficient is also a good thing since this coefficient is the dampening factor expressing how much disturbances are reduced when a bus advances by one station. This clearly shows that cruising guidance makes the system more resilient.

The arguments leading to (9) and (14) expressing the average steady state behavior of a homogeneous line can be repeated and the result is again analogous:

$$\langle \varepsilon_{n,s} \rangle \;\leq\; \frac{\sigma^2 + \langle e \rangle}{1 - (\alpha - \tau)^2} \qquad \text{where} \quad \alpha - \tau \in (0, 1). \tag{17}$$

This expression quantifies how punctuality improves with cruising guidance in the linear regime.

The effect of cruising guidance should also be explored for the non-linear regime. In the interest of brevity, a simulation of idealized systems paralleling the non-linear explorations of Sec. 5 is not provided, since it is reasonable to expect a similar relationship between the linear and non-linear cases now. Instead the system's performance was tested with less broad but more detailed simulations and field tests of a real system. This is explained in the following section.

8. Field Study and Simulation Tests

The experiments were carried out in San Sebastián (Spain) with the cooperation of Dbus, the local bus agency. Dbus is recognized internationally as a provider of good service and a leader in the adoption of novel technologies.[11] The site was a long, busy corridor served by two overlapping lines (numbers 5 and 25); Argote-Cabanero (2014) provides more detail. Line 25 was scheduled with headways of 20 minutes and line 5 with alternating headways of 6 and 8 minutes.

Dbus' goal was to provide the most regular and frequent service possible in the common corridor. To this end, their analysts had constructed a schedule that optimally interleaved both lines to produce 4-6 minute headways in the corridor. Dbus' analysts recognized that to achieve this kind of regularity the buses had to be tightly kept on schedule. Since this level of schedule and headway reliability was difficult to ensure with their existing operational methods – controlling buses at the terminal stations on a schedule basis – management agreed to test the proposed dynamic approach. Dbus purchased consumer Android tablets (Samsung Galaxy Tab 2) and mounted them on

[11] It was one of the first agencies in Spain to achieve public transit standard UNE-EN 13816 for several of its lines.

bus dashboards. VIA Analytics developed an app to implement the proposed control algorithm and archive all the performance data for evaluation. The value of $\bar{z}$ at the last station of the common corridor, where schedule deviations would be greatest, was selected as the evaluation's measure of performance.

Prior to activating the control, Dbus' performance with their original form of static control at the terminal stations was benchmarked from historical data. The dashed horizontal line of Fig. 7 is this benchmark. Departures from their termini where static control was applied were already very reliable – over 95% on time (Vallejo, 2014). This system was also simulated to provide a secondary benchmark. This is shown by the solid horizontal line of the figure. The simulated results are better than the real presumably because real drivers do not follow the control instructions as well as the ideal.[12]

The simulation tool was also used to evaluate the proposed control for different values of α. To this end, rule (11) without cruising guidance, was applied to the two lines in the corridor for 50 simulated days. The average performance is displayed in Fig. 7 by means of a curved solid line, which at its lowest is more than 35% below the horizontal benchmarks. This suggests that the control algorithm might be able to reduce $\bar{z}$ by up to 35% over what is already a very reliable system.

Pursuant to this exercise the control was activated in the real world with a high value of α ($\alpha = 0.9$). This was done for reasons of driver psychology because smaller values imply larger and more variable holding times, which could stress drivers and reduce acceptance. After a period of training, driver compliance turned out to be fairly good; see Fig. 8a. Non-compliance was quantified by the frequency with which each driver violated his/her holding orders by departing too early. The figure includes data from 355 drivers who received at least 100 holding instructions over the course of 5 months. The median driver violates instructions only 13.8% of the time, and the worst driver only by 26%. (The violation values range from -6s to -12s, and are higher for the less compliant drivers.)

The observed reliability level with $\alpha = 0.9$ is shown by the square of Fig. 7. This is about 22% better than the original performance. The square

[12]The simulation tool is trustworthy because its predictions matched Dbus' metrics quite well (Argote-Cabanero, 2014), and prior to this has reproduced successfully the performance of larger systems such as those in Hong Kong and Barcelona, see Via Analytics (2015).

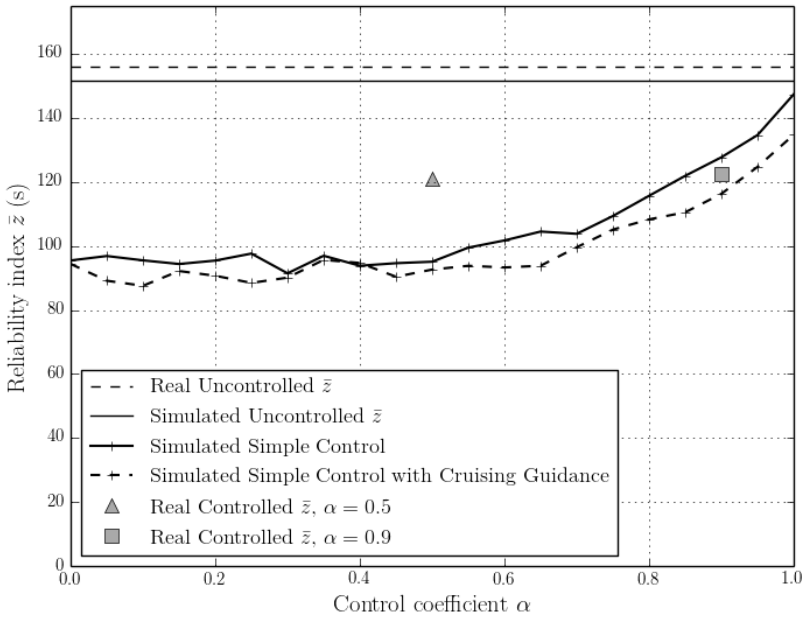
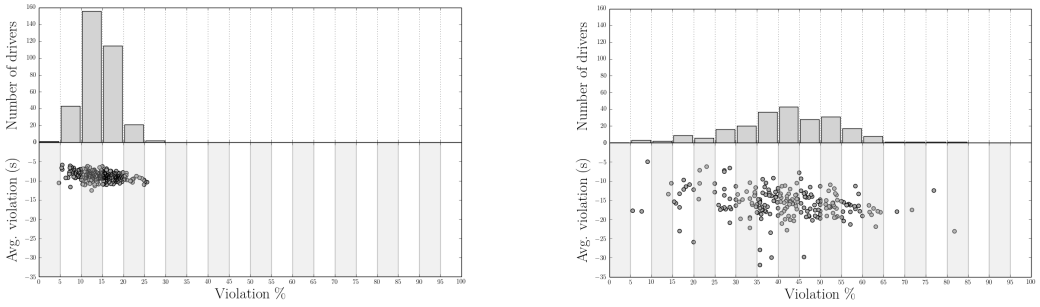


Figure 7: Simulated and observed performance of Dbus' real system.

is even below the solid (simulated) curve. This positioning suggests that Dbus' drivers are doing something above and beyond the ideal drivers of the simulation. The extra improvement is most likely due to cruising guidance because *only* the real drivers receive it – in the form of a continuously updating stimulus-bar that indicates the current schedule deviation as suggested in Sec. 7.[13]

To test the cruising-benefit hypothesis, scatter-plots of the stimulus, $\varepsilon_{n,s}$, vs. the resulting relative difference of the cruising time with the mean, $(c^o_{n,s} - c_{n,s})/c_{n,s}$, were produced. To control for spatio-temporal effects, each plot consists of inter-stop trips outbound from a specific station s in a specific time window of the day. Figure 9 shows the plots for the fifth and sixth stations in the downtown direction in the 4:00-5:00 PM time window over a period of a month. The figure also shows a line fitted to these data, denoted $\hat{c}_{n,s}(\varepsilon_{n,s})$. A

[13]The feature was not simulated a priori because the response of drivers to stimuli could not be calibrated until *after* the system was activated.



(a) $\alpha = 0.9$ (b) $\alpha = 0.5$

Figure 8: Driver compliance results for two deployed α values. Top: violation percentage histogram. Bottom: violation percentage vs average violation magnitude.

similar analysis was performed for all other stations in the two lines. It was found that all statistically significant slopes were negative (as expected) but significance was only achieved where the bus lane was segregated from other vehicular traffic for a significant distance: for 8 out of 23 stations. These negative trends show that drivers respond to cruising stimuli when they can in the way assumed in Sec 7, which was found to improve reliability in the linear regime. Thus, it is reasonable to expect that cruising guidance also improves the reliability of the real (non-linear) system.

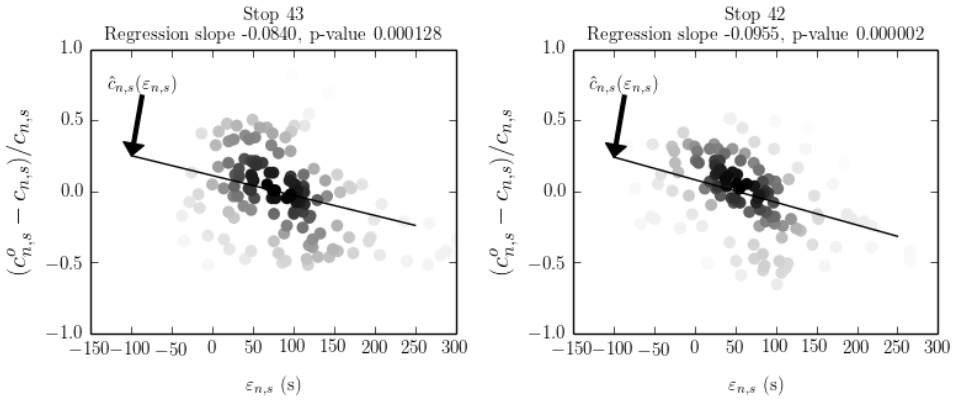


Figure 9: Schedule deviation vs relative cruising time difference for trips outbound from stations 43 and 42, 4:00-5:00pm during October 2014. Darker tones indicate higher measurement density.

To confirm this hypothesis, the ideal-driver simulations were repeated

with cruising guidance a posteriori; i.e., using the dynamic cruising times $\hat{c}_{n,s}(\varepsilon_{n,s})$ estimated from the scatter-plots instead of the static cruising times.[14] The dashed curve at the bottom of Fig. 7 depicts the results. With perfectly compliant drivers, cruising guidance has an effect roughly equal to the sparation between the solid and dashed curves. With real drivers, the improvements should be similar because the degradation in reliability due to driver non-compliance appears to be invariant with control – as Fig. 7 shows, it is roughly the same with cruising guidance (measured by the separation between the square and the dashed curve) and with static control (measured by the separation between the real and simulated benchmarks).

Because Fig. 7 points to an additional benefit by reducing α, a smaller value was also tried ($\alpha = 0.5$). This was done without alerting or retraining the drivers. Perhaps for this reason the results were not as good as expected; see the solid triangle in Fig. 7. This is probably due to driver compliance which turned out to be much lower, as we had feared; see Fig. 8b and compare with Fig. 8a. The new data were collected over a period of three weeks from 224 drivers who received at least 10 holding instructions. The median driver violated holding instructions 42.1% of the time, the best 5% and the worst 80%. (The average magnitude of these violations ranged from -5s to -32s with an average of -15.8s.) Pending further evaluation, the experiment was ended and the value reset to the original, $\alpha = 0.9$.

It should be said in defense of drivers that hardware imperfections can also contribute to discrepancies between real results and the simulations. A system audit during the aforementioned five months revealed that the GPS function of the on-board tablets malfunctioned 1.73% of the time. This must have hindered performance, albeit probably very little. Be that as it may, the improvement in reliability was achieved despite both, imperfect drivers and imperfect hardware. This confirms the resiliency claims of Sec. 6.

9. Concluding thoughts

This paper has shown that it is possible to control dynamically very complex transit systems so they can better stay on schedule. With the proposed control algorithm, reliability can be improved considerably even for systems that have negligible slack as often occurs in practice. Improvements are

[14] This is the same as adding a correction term equal to $\hat{c}_{n,s}(\varepsilon_{n,s}) - c_{n,s}$ to the right hand side of (10).

obtained with both versions of the algorithm – with and without cruising guidance – but the benefits are greater with cruising guidance.

The algorithm reacts in such a simple way to the current state of the system that its application does not increase in complexity with system size. In addition, buses are isolated from one another so that a bus' performance mostly depends on what's happening to it. As a result, the algorithm is resilient to commonplace disruptions such as those due to malfunctioning equipment or non-cooperative drivers.

The algorithm's overall performance does depend on the average driver behavior, however. It was also shown that driver compliance was highly variable. Thus, an area of further research is the identification of triggers that would motivate reluctant drivers to better follow instructions. Other areas deserving of research include the coordination of inter-line transfers, and the coordination of the control with smart forms of traffic signal priority.

Acknowledgments

The authors wish to thank Javier Vallejo (Director of Operations), Eduardo González (Director of Maintenance), Gerardo Lertxundi (General Manager), and the rest of the Dbus team involved in the control implementation. The authors are also grateful to the board of VIA Analytics for their encouragement during this study and to Yiguang (Ethan) Xuan for his contribution to the development of the control app. This research was supported in part by NSF and by UCCONNECT.

References

Argote-Cabanero, J., 2014. Improving Bus Service with Dynamic Holding Control: from Theory to Practice. Ph.D. thesis, University of California, Berkeley.

Bartholdi, J. J., Eisenstein, D. D., 2012. A self-coördinating bus route to resist bus bunching. Transportation Research Part B: Methodological 46 (4), 481–491.

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

Daganzo, C. F., Pilachowski, J., 2011. Reducing bunching with bus-to-bus cooperation. Transportation Research Part B 45 (1), 267–277.

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

Eberlein, X. J., Wilson, N. H. M., Bernstein, D., 2001. The holding problem with real-time information available. Transportation Science 35 (1), 1–18.

Liu, Z., Yan, Y., Qu, X., Zhang, Y., 2013. Bus stop-skipping scheme with random travel time. Transportation Research Part C: Emerging Technologies 35, 46–56.

Lizana, P., Muñoz, J. C., Giesen, R., Delgado, F., 2014. Bus Control Strategy Application: Case Study of Santiago Transit System. Procedia Computer Science 32, 397–404.

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

Osuna, E. E., Newell, G. F., 1972. Control strategies for an idealized public transportation system. Transportation Science 6 (1), 52–72.

Sun, A., Hickman, M., 2005. The real-time stop skipping problem. Transportation Systems 9 (2), 91–109.

Vallejo, J., 2014. personal communication.

Via Analytics, i., 2015. Simulation of real-world bus systems.
URL `www.v-a.io/blog/simulation-of-real-world-bus-systems/`

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

Appendix A. Parameter Estimation

Appendix A.1, below, shows how to estimate $\beta_{n,s}$ and $B_{n,s}$ in (10) and (11) from the data that is harvested during the course of operation. Appendix A.2 then explores how estimation errors affect punctuality.

Appendix A.1. Estimation method

The proposed estimation method assumes that the bus is operated with a control formula (11) based on erroneous initial estimates $\tilde{\beta}_{n,s}$ and $\tilde{B}_{n,s}$. The bus, however, responds according to (10) with the true unknown values. The system therefore obeys a dynamic equation that combines these two versions of (10) and (11).

Consideration shows that this combination can be succinctly expressed as follows:

$$\varepsilon_{n,s+1} = I_{n,s} + \beta_{n,s}\varepsilon_{n,s} - F_{n,s}(B_{n,s}, E_{n,s}) + \nu_{n,s+1}, \tag{A.1}$$

where

$$I_{n,s} = F_{n,s}(\tilde{B}_{n,s}, E_{n,s}) + (\alpha - \tilde{\beta}_{n,s})\varepsilon_{n,s}. \tag{A.2}$$

Note that $I_{n,s}$ is observed. Thus, in (A.1), the variables $\varepsilon_{n,s+1}$, $I_{n,s}$, $\varepsilon_{n,s}$ and $E_{n,s}$ are observed data, the variables $\beta_{n,s}$ and $B_{n,s}$ are unknown parameters, and $\nu_{n,s+1}$ is an unobserved error term. In other words, (A.1) defines a non-linear regression model for the estimation of $\beta_{n,s}$ and $B_{n,s}$, with data spanning a number of days.

In most practical applications the system effect term can be specified to be linear in the parameters, so the regression is linear and one can use consistent BLUE estimators. These can then be used to update the old estimates with standard statistical methods, and update the control. The estimation/updating procedure can be repeated on an ongoing basis to always run the control with a "fresh" set of estimates.

Appendix A.2. Effect of estimation errors on punctuality

Let us now examine the effect of inaccuracy in $\tilde{\beta}_{n,s}$ and $\tilde{B}_{n,s}$ on the system's performance between parameter updates. For this analysis $\tilde{\beta}_{n,s}$ and $\tilde{B}_{n,s}$ are treated as known constants, $\beta_{n,s}$ and $B_{n,s}$ as unknown constants, and $\varepsilon_{n,s}, E_{n,s}$ and $\nu_{n,s+1}$ as random variables that vary from day to day. With this in mind, let us put (A.1) and (A.2) in the following form:

$$\varepsilon_{n,s+1} = \tilde{\alpha}\varepsilon_{n,s} + \tilde{\nu}_{n,s+1}, \tag{A.3}$$

where

$$\tilde{\alpha} = (\alpha + \beta_{n,s} - \tilde{\beta}_{n,s}), \qquad \text{and} \tag{A.4a}$$

$$\tilde{\nu}_{n,s+1} = [F_{n,s}(\tilde{B}_{n,s}, E_{n,s}) - F_{n,s}(B_{n,s}, E_{n,s})] + \nu_{n,s+1}. \tag{A.4b}$$

Note that $\tilde{\alpha}$ is fixed and $\tilde{\nu}_{n,s+1}$ is random. Furthermore, since the control does not affect any of the random variables on the RHS of (A.4b), its LHS ($\tilde{\nu}_{n,s+1}$) can be treated as exogenous noise. The variance of $\tilde{\nu}_{n,s+1}$ shall be denoted $\tilde{\sigma}^2_{n,s+1}$. Note, $\tilde{\sigma}^2_{n,s+1} > \sigma^2_{n,s+1}$.

In most applications the systems' deviations from schedule $E_{n,s}$ should not be significantly correlated with the line's deviations $\varepsilon_{n,s+1}$, so the noise $\tilde{\nu}_{n,s+1}$ should be uncorrelated with the $\varepsilon_{n,s+1}$'s. This means that (A.3) has the same physical interpretation as in the case with known parameters and can be treated recursively as (6) with the same results. In particular, the new version of (9) for a homogeneous line becomes:

$$\text{var}(\varepsilon_{n,s}) = \frac{\tilde{\sigma}^2}{1-\tilde{\alpha}^2} \qquad \text{where} \quad \tilde{\alpha} \in (0,1). \tag{A.5}$$

A comparison of (9) and (A.5) reveals how estimation errors affect system reliability. Note from (A.4) that the estimation error in β determines $\tilde{\alpha}$ and the error in the system parameters, B, determines $\tilde{\sigma}^2$, where $\tilde{\sigma}^2 > \sigma^2$. Thus, the inaccuracy in B has the same effect as an increase in the driver/traffic/passenger noise from σ^2 to $\tilde{\sigma}^2$.

The estimation error in the line's parameter β has a different effect. It affects the choice of the control parameter α through the constraint $\tilde{\alpha} \in (0,1)$. Because the difference $\beta - \tilde{\beta}$ in the homogeneous version of (13) is unknown, to satisfy the constraint we need to choose α assuming that this difference is at the upper end of what is likely, say, two standard deviations of our estimator for β, σ_β. But this is the same as assuming in (A.4a) that the true dimensionless demand β is $2\sigma_\beta$ units larger than in reality, and that there is no estimation error. Thus, the effect on punctuality of not knowing the line's parameter is akin to that of an increase in demand equal to the amount of the uncertainty. The relationship between demand and unpunctuality has been studied extensively in Daganzo (2009) and Xuan et al. (2011).