Title

Reducing bunching with bus-to-bus cooperation

Authors

Daganzo, Carlos F.
Pilachowski, Josh

Publication Date

2009-12-01

Reducing bunching with bus-to-bus cooperation

Daganzo, Carlos F., and Josh Pilachowski

WORKING PAPER
UCB-ITS-VWP-2009-12

UC Berkeley Center for Future Urban Transport
A VOLVO Center of Excellence

December 2009

Reducing bunching with bus-to-bus cooperation

Carlos F. Daganzo and Josh Pilachowski
Institute of Transportation Studies and
Department of Civil and Environmental Engineering
University of California, Berkeley CA 94720

(December 11, 2009)

Abstract

Schedule-based or headway-based control schemes to reduce bus bunching are not resilient because they cannot prevent buses from losing ground to the buses they follow when disruptions increase the gaps separating them beyond a critical value. (Following buses are then overwhelmed with passengers and cannot process their work quick enough to catch up.) This critical gap problem can be avoided, however, if buses at the leading end of such gaps are given information to cooperate with the ones behind by slowing down.

This paper builds on this idea. It proposes an adaptive control scheme that adjusts a bus cruising speed in real-time based on both, its front and rear spacings much as if successive bus pairs were connected by springs. The scheme is shown to yield regular headways with faster bus travel than existing control methods. Its simple and decentralized logic automatically compensates for traffic disruptions and inaccurate bus driver actions. Its hardware and data requirements are minimal.

1. Introduction

It has been known for nearly 50 years (Newell and Potts, 1964) that if buses are left to their own devices, they cannot stay on schedule. The reason is that a lagging bus has to collect more passengers, and therefore tends to fall further behind. And the reverse is also true. So there is a positive feedback loop that leads to undesirable bunching. Transit agencies typically deal with this problem by inserting slack into their schedules, enforcing the latter at control points. Slack, however, reduces the buses' commercial speed - i.e., the average speed that passengers experience including stops. Formulae in Daganzo (1997) quantify this reduction; they show that this medicine (slack) is sometimes worse than the illness (irregular headways).

In view of this, adaptive control schemes based on real-time information have been proposed. They are of three types: simulation-heuristics (see e.g., Hickman, 2001), optimization (e.g., Eberlein et al, 2001) and control (Daganzo, 2009).[1] As explained in this latter reference, the control approach can systematically account for the uncertainties due to traffic and demand while reducing the commercial speed only slightly. Therefore, this paper will be in the control genre, and build on this last reference.

This reference, and the present paper, propose control strategies for high-frequency bus or transit lines run without a schedule. Their goal is providing quasi-regular headways while maintaining as high a commercial speed as possible. Daganzo (2009) proposes to hold buses at discrete control points for brief periods of time that depend on the time interval since the passage of the previous bus - the headway. Although the method turns out to be quite efficient when disturbances are small, its forward-looking/non-cooperative character (buses are not allowed to slow down in response to events behind) hinders performance whenever disturbances are large. In fact, the method cannot prevent collapse if disturbances create supra-critical gaps in the sense of Newell (1977).

To alleviate this problem, a cooperative, two-way-looking strategy based on the spacings in the front and back of each bus is proposed and evaluated here. To further enhance performance, the strategy will allow these spacings to be monitored as frequently as desired. The paper is organized as follows. Section 2 examines an idealized version of the problem for which cooperative strategies can be easily designed. Section 3 proposes one such strategy. Section 4 develops approximate performance formulae for the idealized scenario and compares the proposed strategy with non-cooperative counterparts. Section 5 generalizes the results to real settings; it is found that real systems are more easily controlled than idealized ones and that real performance is well predicted with the idealized formulae. Finally, Section 6 discusses potential improvements and future work.

[1] There is also a literature dealing with small systems and strategies that do not fully use available real-time information; e.g., Osuna and Newell (1972), Barnett (1974), Newell (1974), Ignall and Kolesar (1974), Hickman (2001) and Zhao et al. (2006).

2. A continuum idealization

Considered is a closed loop of length L (km) on which $n = 1, 2...N$ transit vehicles operate. These transit vehicles shall be called "buses". The loop includes $m = 1, 2, ...M$ stops with an even inter-stop distance, $D = L/M$. Passengers arrive along this loop as a spatially homogeneous, time-independent process with independent increments, average rate λ (pax/km-hr) and index of dispersion γ (pax). They materialize uniformly along the route but board the bus at the nearest stop, so the average demand rate at each stop is $\lambda L/M$ (pax/hr). Buses are assumed to cruise at an average speed of $\bar{v}$ (km/hr). This is the maximum speed that they can sustain including random delays due to traffic. Their average commercial speed is lower because buses are delayed by serving passengers.

This model is now simplified so it can be used to identify control laws and quantify their performance. It shall be assumed that bus delays due to passenger-related stops are directly proportional to the number of passengers boarding the bus, and that these delays are continuously distributed along the perimeter of the loop as if passengers were infinitely divisible and were picked up where they appeared. This is why the approximation is labeled "continuum". In this continuum model bus travel time has two components: a part that is independent of the number of passengers and inversely proportional to the cruising speed, and a part that accrues with the passengers the bus meets at a rate of b (hr/pax).

2.1. Equilibrium relations

Consider now an equilibrium where passengers arrive at a steady rate, buses are evenly separated with spacing $S = L/N$, and travel with the same commercial speed, $\bar{c}$. In this equilibrium the headway, H, is:

$$H = S/\bar{c}. \tag{1}$$

A relation between the equilibrium $\bar{c}$ and $\bar{v}$ is obtained by decomposing the time it takes a bus to travel one distance unit, $1/\bar{c}$, into time cruising, $1/\bar{v}$, and time collecting passengers, $b\lambda H = b\lambda S/\bar{c}$. The resulting equality, $1/\bar{c} = 1/\bar{v} + b\lambda S/\bar{c}$, can be rearranged to yield:

$$\bar{c} = \bar{v}(1 - \lambda b S). \tag{2}$$

2.2. Quasi-equilibrium, stochastic law of motion

Consider now the system's stochastic behavior when all its variables are close to equilibrium. Sought is an expression for the position of vehicle n at time $t + \Delta t$ in terms of its position at time t.

Let "y_n" denote the position of bus n at some t, and "$s_n = y_{n\ominus 1} \ominus y_n$" the spacing in front of it. The "o-minus" notation denotes substraction modulo N for the vehicle number subscript, and modulo L for the vehicle positions. (Recall, we have N vehicles; thus, when applied to vehicles,"o-minus" denotes the vehicle in front; and "o-plus" shall denote the vehicle behind.) In order to eliminate the time variable from the notation the comparison operator ":=" will be used. It will signify that the values on its LHS are evaluated one step Δt after those on its RHS. For example, if vehicle n has commercial speed c'_n at time t, then:

$$y_n := y_n \oplus c'_n \Delta t. \tag{3}$$

The relation between c'_n and the current cruising speed $v'_n \leq \bar{v}$ is examined now.

Assume for the moment that the demand is deterministic and consider what happens to the bus in one time step Δt, while it travels a distance $c'_n \Delta t$. The number of passengers it collects is the product of the demand rate, the headway and the distance traveled. Since the system is in quasi-equilibrium, this product is approximately $\lambda(s_n/c'_{n\ominus 1})(c'_n \Delta t)$. Now, since both commercial speeds should be close to the average, and there is no reason for one to be greater than the other, the number of passengers can be approximated by $\lambda s_n \Delta t$. Therefore the bus is stopped for a time $\lambda b s_n \Delta t$, leaving $(1 - \lambda b s_n)\Delta t$ time units available for cruising. The cruising distance is therefore approximately given by $v'_n(1 - \lambda b s_n)\Delta t$. Since the cruising distance must equal $c'_n \Delta t$, it follows that:

$$c'_n \approx v'_n(1 - \lambda b s_n). \tag{4}$$

Note how for any given v'_n lower cruising speeds arise as a result of greater spacings. Thus, if no control was exercised and the v'_n were allowed to remain constant, buses that fall behind would tend to fall further behind and those that catch up would speed up. They would bunch. As shown below, this can be avoided by adjusting v'_n dynamically.

3. A control law for the idealized scenario

A quasi-equlibrium, stochastic law of motion is obtained by combining (3) and (4), and adding noise terms, ν_n, to capture the effect of traffic disturbances, driver errors and randomness in passenger arrivals:

$$y_n :\approx y_n \oplus c'_n \Delta t \oplus \nu_n \approx y_n \oplus v'_n(1 - \lambda b s_n)\Delta t \oplus \nu_n, \qquad n = 1, 2, ...N. \tag{5}$$

Cruising speeds are dynamically adjusted as a function of the spacings as per:

$$v'_n \doteq \bar{v} + C_n(\boldsymbol{s}, S), \qquad n = 1, 2, ...N, \tag{6}$$

where $\boldsymbol{s}$ the vector of spacings and $C_n(\boldsymbol{s}, S)$ is:

$$C_n(\boldsymbol{s}, S) \doteq [\lambda b \bar{v}(s_n - S) + \alpha(s_n - s_{n\oplus 1}) - \delta]/(1 - \lambda b s_n), \qquad n = 1, 2, ...N. \tag{7}$$

The constants α and δ are free parameters.

It will be convenient to denote the numerator of $C_n(\boldsymbol{s}, S)$ by F. The three terms of F attempt three things: (i) accelerating a bus when its spacing is greater than the equilibrium target; (ii) retarding it if the following bus is far behind; and (iii) ensuring that $F \leq 0$ so that target cruising speed does not exceed $\bar{v}$. The control law is two-way-looking because a bus reacts to both, its front and rear spacings.

The special form (6-7) has been chosen because, as is now shown, once inserted in (5) it produces a stochastic law of motion in the family studied in Daganzo (2009), which overcomes the bunching problem. To see how this comes about insert (7) in (6), and multiply both sides of the result by $(1-\lambda b s_n)$. Then, using (2) and (4) to express the result in terms of commercial speeds, one finds: $c'_n \approx \bar{c} - \lambda b \bar{v}(s_n - S) + [\lambda b \bar{v}(s_n - S) + \alpha(s_n - s_{n\oplus 1}) - \delta]$. Since the two λ-terms cancel out, the result is:

$$c'_n \approx \bar{c} + \alpha(s_n - s_{n\oplus 1}) - \delta. \tag{8}$$

This shows that the equilibrium speed achieved with control is $\bar{c} - \delta$; i.e., that it is δ units lower than the theoretical maximum, $\bar{c}$. Now,use $\bar{c}' = \bar{c} - \delta$ for the equilibrium speed with control, and combine (5) and (8) to write:

$$y_n :\approx y_n \oplus \bar{c}'\Delta t \oplus \alpha(s_n - s_{n\oplus 1})\Delta t \oplus \nu_n, \qquad n = 1, 2, ...N. \tag{9}$$

Next, express (9) in terms of deviations, ϵ_n, from the equlibrium trajectories, x_n, which are:

$$x_n := x_n \oplus \bar{c}'\Delta t, \qquad n = 1, 2, ...N. \tag{10}$$

To introduce the deviations, subtract (10) from (9) and make the substitutions: $s_n = S + \epsilon_{n\ominus 1} - \epsilon_n$ and $s_{n\oplus 1} = S + \epsilon_n - \epsilon_{n\oplus 1}$. The result is:

$$\epsilon_n :\approx \epsilon_n \oplus (\alpha\Delta t)\epsilon_{n\ominus 1} \oplus (1 - 2\alpha\Delta t)\epsilon_n \oplus (\alpha\Delta t)\epsilon_{n\oplus 1} \oplus \nu_n, \qquad n = 1, 2, ...N. \tag{11}$$

Except for the modular arithmetic, due to the closed loop, this system of linear difference equations with random impulses is identical to the open-loop systems analyzed in Daganzo (2009). Therefore, if system (11) dissipates disturbances over small distances compared with L, the closedness of the loop should not come into play and the results in Daganzo (2009) should apply approximately. This should happen when the number of buses on the loop is large compared with 1. This fact is used below to evaluate the control law.

4. Performance

Proposition 2 in Daganzo (2009) states that if the ν_n are uncorrelated with identical variances (u.i.v.), and $\alpha\Delta t \in (0, 0.5)$ so that the coefficients of (11) from a pdf, then the ratio of $var(s_n)$ (the mean squared error of the infinite time-series $\{s_n\}$) and $var(\nu_n)$, is a constant that only depends on the coefficients of (11). This suggests that, in our case, $var(s_n)/var(\nu_n)$ should also be a constant that only depends on $\alpha\Delta t$ and N. Appendix A proves that this is in fact true.[2]

[2]Appendix A also shows that if the ν_n are not u.i.v., but are uniformly bounded for all t by a quantity $\bar{\nu}$, then the deviations $\{s_n - S\}$ are uniformly bounded by a quantity that only depends on $\bar{\nu}$, N and $\alpha\Delta t$, provided still that $\alpha\Delta t \in (0, 0.5)$.

4.1. Approximate formulae

Simulations in Daganzo (2009) under the u.i.v. assumption also show that $var(s_n)/var(\nu_n)$ roughly equals the variance of the coefficients of the dynamic equation; i.e., (11) in our case. Therefore, it is conjectured (and later demonstrated with simulations) that this is also approximately true in our case, if the system includes more than just a few buses.[3] Note, under the u.i.v. assumption, $var(\nu_n)$ should be of the form: $var(\nu_n) = r^2\Delta t$, where r^2 is the variance of the noise per unit time (km^2/hr). Since the variance of the coefficients of (11) is $2\alpha\Delta t$, the following approximation should hold for problems with many buses:

$$var(s_n) \approx \tfrac{1}{2}r^2/\alpha, \qquad n = 1, 2, ...N. \tag{12}$$

This result applies to the infinite time series of spacings. However, if a system is observed for a short time, the initial transient will not have been dissipated and the variance of the spacings will depend on the initial conditions. To estimate the transient, note that if we let $\Delta t \to 0$ in (11) the only time constant that can be formed from the data defining the problem is the parameter $1/\alpha$. Hence, the transient should be comparable with this quantity. As a result, the observation time required for (12) to apply should be several times larger than $1/\alpha$.

Equation (12) can be used to choose α and δ that optimize the control. For example, assume that the transit agency wishes to maximize the buses' commercial speed (i.e. minimize δ) while preventing bunching. If the sytem is to behave as predicted, the quantity $F = [\lambda b\bar{v}(s_n - S) + \alpha(s_n - s_{n\oplus 1}) - \delta]$ should stay negative most of the time. Hence, δ is constrained to equal or exceed 3 standard deviations of $G \doteq \lambda b\bar{v}(s_n - S) + \alpha(s_n - s_{n\oplus 1})$. This standard deviation depends on the correlation between s_n and $s_{n\oplus 1}$. Since these correlations should be slightly negative, the formulas below assume that the correlation is -0.25.[4] The reader can verify that $var(G) \approx [2.5\alpha^2 + 2.5\lambda b\bar{v}\alpha + (\lambda b\bar{v})^2]var(s_n)$; and using (12), that:

$$var(G) \approx \tfrac{1}{2}r^2[2.5\alpha + 2.5\lambda b\bar{v} + (\lambda b\bar{v})^2/\alpha]. \tag{13}$$

Thus, with the criterion $\delta \approx 3\sqrt{var(G)}$, the smallest feasible δ for a given α is:

$$\delta \approx 2.12r[2.5\alpha + 2.5\lambda b\bar{v} + (\lambda b\bar{v})^2/\alpha]^{1/2}. \tag{14}$$

The value of α that minimizes this expression is:

$$\alpha^* \approx 0.63\lambda b\bar{v}; \tag{15}$$

and the corresponding reduction in average commercial speed is therefore:

$$\delta^* \approx 5.0r(\lambda b\bar{v})^{1/2}. \tag{16}$$

[3]Appendix A gives an exact formula for $var(s_n)$, which is too complicated to be useful.

[4]This value produces results within 16% of the worst case with correlation -1, and smaller errors for the optimistic case with zero correlation. So the approximation is robust.

Finally, the variance of the spacing under this type of control is:

$$var(s_n^*) \approx 0.79r^2/(\lambda b\bar{v}). \tag{17}$$

4.2. Comparison with one-way-looking

Appendix B rederives the above expressions for a forward-looking control law which includes only s_n as a state variable. It shows that if this law is applied in continuous time (i.e., with $\Delta t \to 0$) then the commercial speed penalty (16) increases by 20% and the spacing variance (17) by 25%. So, two-way looking is not just good because it overcomes the critical gap problem, it also performs better under ordinary conditions.

The headway control strategy in Daganzo (2009) should perform similarly as the strategy of Appendix B since both are forward-looking – or perhaps a little worse because headway control is not applied continuously but at discrete points along the route. For the following data ($\lambda = 27$ pax/hr-km, $b = 4$ s, $\bar{v} =$ 20 km/hr, $r^2 = 0.12$ km^2/hr (which corresponds to a standard error of about 0.045 km per minute) the formulae just presented yield: $\alpha^* = 0.38$ hr^{-1}, δ^* $= 1.3$ km/hr and $\sqrt{var(s_n)} = 0.40$ km. If buses are closely spaced with say $S = 1.5$ km, then: $\bar{c} = 19.1$ km/hr, $\bar{c}' = 17.8$ km/hr, and $H \approx 5.0$ min. For these commercial speeds, the standard deviation in the headways is about 1.5 min and the increase in in-vehicle travel time due to the control is about 13.8 s per km traveled. The time added by the headway-based control strategy in Daganzo (2009) for a similar set of conditions is 19 s/km (37% greater); and the time added by conventional schedule control about 60 s/km (about 334% greater). Thus, the proposed strategy shows promise.

The results in this section are approximate, however, and only pertain to the idealized model. Therefore, the next section generalizes the model and the control law. It shows that real systems perform approximately as predicted in this section – slightly better in fact.

5. Generalized results for realistic settings

This section proposes and evaluates control laws for realistic settings with discrete stops and passengers, where all passengers access the bus to/from these stops. It is assumed that buses do not skip stops and that alighting passengers do not contribute strongly to the bus dwell time at a stop. More specifically, buses are delayed by a fixed amount, τ, each time they pass a stop (corresponding to the time it takes the bus to decelerate, accelerate open and close its doors), and an additional time B for each boarding passenger. The control laws about to be derived can also be used without these assumptions; but these are the assumptions under which they will be tested.

Noise in a discrete system arises both because of traffic/driver errors, and because of variations in the number of passengers that board at each stop. As before, the traffic/driver noise is assumed to obey a stationary stochastic process with independent increments. Its variance rate is denoted R_T^2 (km^2/hr). Passengers are assumed to arrive at the stops according to stochastic processes with independent increments and index of dispersion γ (p). Thus, the quai-equilibrium variance rate of the time added at each stop is approximately: $\lambda\gamma DHB^2$ (hr^2). Therefore, the variance rate per unit distance is $\lambda\gamma HB^2$ (hr^2/km). Since vehicles travel at an average speed of $\bar{c}'$ the variance rate of the distance traveled per unit time is approximately: $\lambda\gamma HB^2\bar{c}'^3 = \lambda\gamma SB^2\bar{c}'^2$ (km^2/hr). The total rate combines traffic and passenger effects. It is: $R^2 \doteq R_T^2 + \lambda\gamma SB^2\bar{c}'^2$ (km^2/hr).

5.1. Three generalized control methods

It is conjectured that a discrete model with $\tau = 0$ will behave similarly as a continuum model with $r = R$ since the effect of an individual passenger is usually small, and its precise location should not matter. (This will be discussed in more detail in the next subsection.) However, because τ is usually at least an order of magnitude greater than B, its lumpiness could have a more significant effect. Therefore, three control methods are now introduced to deal with the lumpiness in τ. All three methods work by introducing an imaginary problem with $\tau^* = 0$ that is similar in behavior to the original problem.

Method 1: The first and most complex method captures the effect of the intervening stops by introducing their fixed delays as state variables. Let the cumulative number of stops visited by bus n be z_n. These quantities are initialized so that the modulo M difference, $z_{n\ominus 1} \ominus z_n$, is at all times the number of stops between buses $z_{n\ominus 1}$ and z_n. The quantities are allowed to be fractional, to account for buses that are at a stop but have not yet been fully delayed. Define then the effective spacings as:

$$s_n^* \doteq s_n + \tau(z_{n\ominus 1} \ominus z_n)\bar{v}. \tag{18}$$

Consider now an imaginary system without fixed stop times, but including an extra distance $\tau\bar{v}$ in front of each stop. Without control, this system should have the same headways and travel times as the original, and its spacings should be related to the original by (18); i.e., the conventional spacings of the imaginary system equal the effective spacings of the real system. Since the two systems are dual images of each other, a dynamic control law such as (6-7) that works well in this imaginary system will also work well in the original. Accordingly, the proposed control law is:

$$v_n' \doteq \bar{v} + C_n(\boldsymbol{s}^*, S^*), \qquad n = 1, 2, ...N, \tag{19}$$

where S^* is the effective equlibrium spacing; i.e., the equlibrium spacing in the imaginary model. The formula is: $S^* \doteq (L + M\tau\bar{v})/N$.

Note every quantity on the RHS of (19) is known at the moment control is applied. Thus, the strategy is feasible. The strategy can be applied with little modification even if the fixed stop times differ across stops. This may be important when one or more of the stops include long periods for drivers to rest.

Method 2: If $D << S$ (i.e., consecutive buses are always separated by many stops) then the benefit of the expanded state variables introduced with Method

1 is diluted. In this case it could be almost as effective, and simpler, to work only with the original spacings. With this method, the fixed stop delays τ would be captured in the imaginary system by reducing its bus cruising speed just enough to preserve the equlibrium bus trip times. The real and imaginary systems are not exact duals of each other but should behave similarly. If we use an asterisk for the vehicle speeds of the imaginary model, the two cruising speeds are related by $1/v_n^* \approx 1/v_n' + \tau/D$. Therefore, differential changes in these speeds are related by $\Delta v_n' \approx \Delta v_n^*(v_n'^2/v_n^{*2}) \approx \Delta v_n^*(1+\tau v_n'/D)^2$. In quasi-equilibrium, where $v_n^* - \bar{v}^* \doteq \Delta v_n^*$ and $v_n' - \bar{v} \doteq \Delta v_n'$, the adjustment in speed for the imaginary model stipulated by (6-7), $v_n^* - \bar{v}^* = C_n(\boldsymbol{s}, S)$, should be small and satisfy approximately the differential relation; i.e., $(v_n' - \bar{v}) \approx (v_n^* - \bar{v}^*)(1 + \tau v_n'/D)^2 \approx C_n(\boldsymbol{s}, S)(1+\tau v_n'/D)^2 \approx C_n(\boldsymbol{s}, S)(1+\tau\bar{v}/D)^2$. Thus, it is proposed to use:

$$v_n' \doteq \bar{v} + (1+\tau\bar{v}/D)^2 C_n(\boldsymbol{s}, S), \qquad n = 1, 2, ...N. \tag{20}$$

Method 3: An alternative approach, still with $\boldsymbol{s}$ as the state (used in Pilachowski, 2009), uses a continuum imaginary model with $\tau^* = 0$ and the same cruising speed at the original, but adjusts the boarding time per passenger, b^*, so as to keep the bus stop times at equlibrium invariant; i.e, so that: $b^*(\lambda HD) = b(\lambda HD) + \tau$. The imaginary boarding time per passenger is therefore: $b^* = b + \tau/(\lambda HD) = b + \tau\bar{c}'/(\lambda SD) \approx b + \tau\bar{c}/(\lambda SD)$. This parameter is then used instead of b in (6-7) to obtain the target cruising speed.

The imaginary system of Method 3 (unlike those of the other 2 methods) assumes that buses bunch even if $b = 0$. As a result, the method could sometimes over-control and reduce the cruising speeds more than necessary. The method is well suited, however, for applications where buses skip stops because then the time buses lose by accelerating/decelerating and opening/closing doors does depend on the number of passengers. In this type of application the marginal time added by a passenger, b^*, could be estimated empirically or analytically. Note however that all three methods are roughly equivalent if $\tau \approx 0$.

5.2. Evaluation

It is claimed that a real system with discrete stops is more easily controlled than its continuum counterpart; i.e., that the predictions of (12) and (14) for the continuum version of a real system are upper bounds to the performance of the latter.

To see why this claim is reasonable consider an extreme case consisting of a very long route with no traffic noise, $\tau = 0$ and many more buses than stops, $M/N >> 1$. Examine now what happens when the separation between bus stops is increased by reducing N while keeping everything else constant, including the control law. First note that the set of buses traveling between stops are deterministically governed by versions of (6-7) and (11) without the noise terms, and that this deterministic control relaxes the bus positions and spacings toward equilibrium exponentially. This means that the distance buses require to (nearly) reach their equilibrium trajectories after experiencing a disturbance at a stop should grows logarithmically with the size of the disturbance. Now, since the demand density is constant, these disturbances should be proportional to the square root of the distance bettween stops. As a result, the distance buses require to reach equilibrium increases by a fixed quantity when the station spacing is doubled; i.e., the road section downstream of a stop where buses relax toward equilibrium becomes a smaller fraction of the inter-stop distance as the latter increases. In other words, the more separated the stations (and the lumpier the disequilibrium impulses) the longer buses stay near equilibrium; i.e., the less the standard deviation of the bus spacings and the headways. This is what was claimed.

Simulation results: Pilachowski (2009) describes a simulation tool for the discrete system. This simulation keeps track of individual passengers at the stops and in the buses, and models the motion of the buses as described in this paper. The only significant difference is that the simulation allows buses to skip stops when nobody needs to board or alight.

Simulation runs confirm the controlability claim. They show that if no control is applied, the continuum problem collapses into bunches before the discrete problem; see Fig. 1.

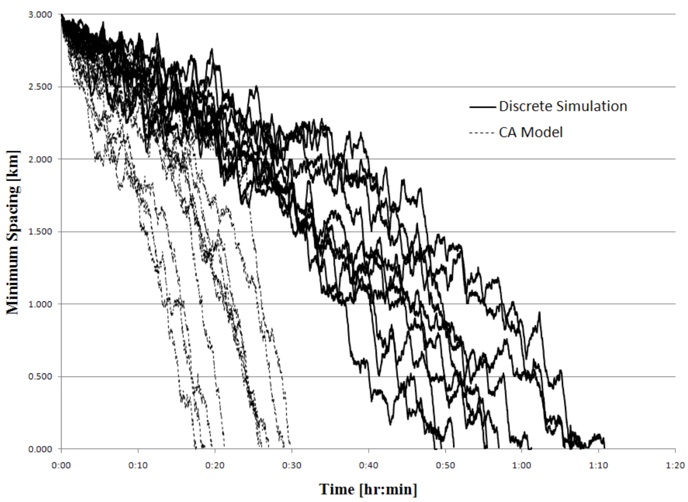

Figure 1: Sample paths of the minimum spacing across all buses assuming, (i) discrete stops (solid lines) and (ii) continuous boarding (dotted lines).

The simulation can also be applied with control methods 2 or 3 by properly selecting its parameters. They fall in 4 categories: bus-related $(N, \bar{v}, \tau, B)$; infrastructure-related (S, D, R_T^2); demand-related (λ, γ); and control-related (α, δ, Δt). The simulation keeps track of the positions and spacings of the buses with a time resolution of 1 s, and returns the variance of the spacings and the headways. Go to: `http://www.ce.berkeley.edu/~daganzo/Simulations/Bus_Bunching.html` for a visualization.

To test the accuracy of the formulas of Section 3, a set of 200 simulations was run, with inputs selected at random from Table 1 and each lasting 8-hrs. Control Method 2 was used.

The results turn out not to be sensitive to Δt, as in the formulas. It is found, also as expected, that the spacings produced by the simulation are slightly more regular than those given by (12) and (14); see the summary results in Figures 2 (for $\tau = 0$) and 3 (for $\tau = 30$s). The figures confirm that cooperative control indeed succeeeds in preventing bunching, and that the predictions of Section 3 are sufficiently accurate for planning purposes.

Table 1: Simulation Data

parameter	range
N	[3, 20]
$\bar{v}$ (km/hr)	[25, 60]
τ (s)	0 or 30
B (s/p)	2 or 4
S (km)	[2, 6]
D/S	2, 4 or 8
R_T^2 (km^2/hr)	0 , 0.1, or 0.4
λ (p/km-hr)	[10, 100]
γ (p)	1
$\alpha/(\lambda B \bar{v})$	0.5, 1 or 2
δ (km/hr)	given by eq.(14)
Δt (s)	5 or 20

6. Discussion: Further work

Non-linear effects: This paper recommends choosing a large enough speed reduction parameter, δ, to ensure that the (linear) control law (6) rarely stipulates speeds greater than $\bar{v}$. However, lower δ's can be used if one is prepared to allow the non-linear relation $v'_n \doteq \text{middle}\{0, \bar{v} + C_n(\boldsymbol{s}, S), \bar{v}\}$ to occur. This is easy to do in practice (and in simulation) but hard to analyze mathematically. Allowing the system to enter the non-linear regime can be beneficial. Simulations show that by reducing δ toward 0, the average commercial speed increases, albeit the headway variance also increases. Simulations also show that, even if δ is chosen as recommended in this paper, the non-linear regime arises when the system is disrupted by a disturbance that creates a sufficiently large supra-critical gap. The simulations also show that the proposed control law can successfully restore order when supra-critical gaps arise. As such, it seems to solve the pesky critical gap problem identified in the introduction. Clearly, a better understanding of the system's behavior under non-linear control is desirable.

More complex scenarios: Although adjustments can be made to the control laws and formulas of this paper to account for time-dependent and space-dependent phenomena (e.g., involving demand and cruising speeds), a systematic understanding of these inhomogeneities is desirable. The ideas in this paper

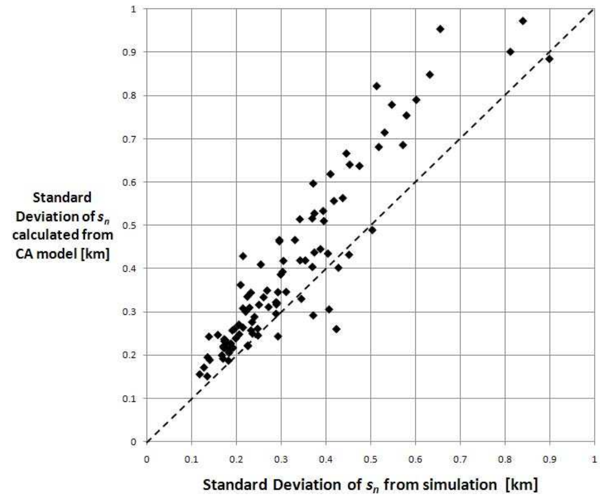

Figure 2: Predicted vs. observed standard deviations of the spacings ($\tau = 0s$)

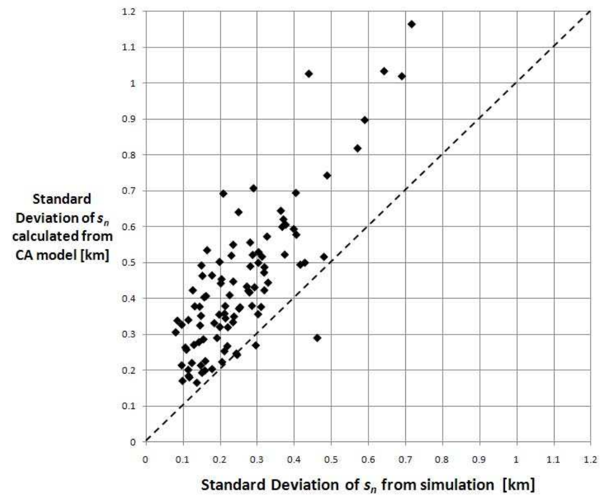

Figure 3: Predicted vs. observed standard deviations of the spacings ($\tau = 30s$)

can also be generalized to systems operated with a schedule, and to corridors operating multiple lines that interact through the sharing of passengers, or through the sharing of a no-passing right-of-way.

Human factors and other practical considerations: Since the inspection and control interval Δt does not influence the results significantly, drivers do not have to monitor their speeds continuously. (Recall that driver errors are accounted for in the control.) Thus, an implementation can succeed by giving drivers rough rules and simple, easy-to-understand, feedback mechanisms. One possibility is a screen that would display a bus' current "speed limit". Drivers would only have to obey this artificial speed limit approximately. Another possibility is remote control of the buses' top speed, which would allow drivers to operate normally. This could be safe for systems with an exclusive right-of-way. Although many implementation details remain to be worked out[5] the proposed strategy can be implemented with current technology. It only requires a GPS-enabled computer on each bus, linked to the control center. Details should be evaluated, not just theoretically, but in the field.

Acknowledgment

Research supported by the VREF International Center of Excellence on Future Urban Transportation at the University of California, Berkeley.

Appendix A. The distribution of spacings

Let $\epsilon_{i,n}$ be the position error of bus n at time $t = i\Delta t$, and $\xi_{i,n} = \epsilon_{i\ominus 1,n} - \epsilon_{i,n}$ the error in the spacing in front of bus n at the same time. This appendix examines the statistical distribution of these variables. To this end, it will be convenient to denote the $(N \times 1)$ column vectors of position and spacing errors at instant i by $\boldsymbol{\epsilon_i}$ and $\boldsymbol{\xi_i}$. An exact formula for $cov(\boldsymbol{\xi_i})$ under u.i.v. noise is given. The formula reveals that the variance of the spacings at any time i, $\text{var}(s_{i,n}) \equiv \text{var}(\xi_{i,n})$ is bounded by a quantity that is independent of both, i and L. It is also shown that if the noise terms are bounded then the $\xi_{i,n}$ themselves are bounded, even if the noise terms are not u.i.v.

To start, rewrite (11) in matrix form as:

$$\boldsymbol{\epsilon_i} = \boldsymbol{A}\boldsymbol{\epsilon_{i-1}} + \boldsymbol{I}\boldsymbol{\nu_{i-1}}. \tag{21}$$

In this expression $\boldsymbol{I}$ is the (N×N) identity matrix; and $\boldsymbol{A}$ is a symmetric, circulant square matrix with elements: $A_{m,n} = 1 - 2\alpha\Delta t$ if $m = n$, $A_{m,n} = \alpha\Delta t$ if $m \ominus n = \pm 1$, and $A_{m,n} = 0$ otherwise.

Now, iterate (21) i times, expressing each time the $\boldsymbol{\epsilon}$ on the right hand side in terms of the $\boldsymbol{\epsilon}$ in the prior time step, to obtain:

$$\boldsymbol{\epsilon_i} = \sum_{j=1}^{i-1} \boldsymbol{A^j}\boldsymbol{\nu_j} + \boldsymbol{A^i}\boldsymbol{\epsilon_0}. \tag{22}$$

Next, write the deviations from the spacings $\boldsymbol{\xi_i}$ using the linear transformation $\boldsymbol{\xi_i} = \boldsymbol{D}\boldsymbol{\epsilon_i}$, where $\boldsymbol{D}$ is the matrix that subtracts the terms of $\boldsymbol{\epsilon_i}$ as per the relation: $\xi_{i,n} = \epsilon_{i,n\ominus 1} - \epsilon_{i,n}$. (The elements of $\boldsymbol{D}$ are: $D_{m,n} = -1$ if $m = n$; and $D_{m,n} = 1$ if $n = m \ominus 1$.) The result is:

[5]For example, for systems operated in traffic one needs to correlate the maximum cruising speed given by (6), which includes traffic disturbances, with the bus's top speed.

$$\boldsymbol{\xi_i} = \boldsymbol{D\epsilon_i} = \sum_{j=1}^{i-1} \boldsymbol{DA^j\nu_j} + \boldsymbol{DA^i\epsilon_0}. \tag{23}$$

This expression involves powers of $\boldsymbol{A}$ and is now simplified. To this end, define the uniform square matrix $\boldsymbol{U}$, with elements $1/N$. The following is true.

LEMMA 1: *If* $0 < \alpha\Delta t < 0.5$, *there is a set of matrices,* $\{\boldsymbol{E_k} : k = 1, 2, ...N\}$, such that:

$$\boldsymbol{A^j} = \boldsymbol{U} + \sum_{k=2}^{N} \lambda_k^j \boldsymbol{E_k}, \tag{24}$$

for a set of constants, λ_k, *such that* $|\lambda_k| < 1$ *for* $k = 2, 3, ...N$.

Proof: If $0 < \alpha\Delta t < 0.5$, then $\boldsymbol{A}$ has the structure of the one-step transition probability matrix of a Markov chain. The chain is obviously non-periodic and communicating; i.e., it is irreducible. Because of this property, the largest eigenvalue of $\boldsymbol{A}$ is $\lambda_1 = 1$; and the rest must satisfy $|\lambda_k| < 1$ (for $k > 1$). So, raising the diagonalized version of $\boldsymbol{A}$ to the the $j^{\text{ th}}$ power yields: $\boldsymbol{A^j} = \sum_k \lambda_k^j \boldsymbol{E_k}$. Now note: for $j \to \infty$, $\boldsymbol{A^j} = \boldsymbol{E_1}$. Because the chain is irreducible, the rows of this matrix must be identical and equal to the steady state probability vector. Furthermore, because $\boldsymbol{A}$ is symmetric, the steady state probability vector has identical elements, $1/N$. Thus, $\boldsymbol{E_1} = \boldsymbol{U}$, and $\boldsymbol{A^j} = \boldsymbol{U} + \sum_{k>1} \lambda_k^j \boldsymbol{E_k}$. QED

LEMMA 2: *The eigenvalues,* λ_k, *and the elements of* $\boldsymbol{E_k} \doteq (e_{kij})$ *are*:

$$\lambda_k = 1 - 2(\alpha\Delta t)[1 - cos(2\pi(k-1)/N)], ; \quad \text{and} \tag{25}$$

$$e_{kij} = cos(2\pi(k-1)(i-j)/N)/N. \tag{26}$$

Proof: Let $\boldsymbol{a}$ be the first column of circulant matrix $\boldsymbol{A}$ and $\boldsymbol{\lambda}$ the vector of eigenvalues. It is well known that the eigenvectors of any circulant matrix are the columns of the $(N \times N)$ discrete Fourier transform matrix, $\boldsymbol{F}$, and that the vector of eigenvalues is given by the product of $\boldsymbol{F}$ and the first column of the circulant matrix. Thus, $\boldsymbol{\lambda} = \boldsymbol{Fa}$ in our case. Inserting $\boldsymbol{a}^T = (1 - 2\alpha\Delta t, \alpha\Delta t, 0, ..., 0, \alpha\Delta t)$ in this expression yields (25). Now, recall that $\boldsymbol{F}^{-1} = \widetilde{\boldsymbol{F}}/N$, where the tilde denotes complex conjugation. Hence $\boldsymbol{A} = \boldsymbol{F}^{-1}\text{diag}(\boldsymbol{\lambda})\boldsymbol{F} = \widetilde{\boldsymbol{F}}\text{diag}(\boldsymbol{\lambda})\boldsymbol{F}/N$, and $\boldsymbol{A}^j = \widetilde{\boldsymbol{F}}\text{diag}(\boldsymbol{\lambda})^j\boldsymbol{F}/N$. Thus, if we denote by $\boldsymbol{F}_k$ the square matrix obtained by multiplying the k^{th} column of $\widetilde{\boldsymbol{F}}$ and the k^{th} row of $\boldsymbol{F}/N$, the last equality can be rewritten as: $\boldsymbol{A}^j = \sum_k \lambda^j \boldsymbol{F}_k$. The eigenvalues and the LHS of this expression are real but the $\boldsymbol{F}_k$ are complex. Thus, conjugation shows that $\boldsymbol{A}^j = \sum_k \lambda^j \widetilde{\boldsymbol{F}}_k$. Now add the last two equalities and divide the result by 2 to find $\boldsymbol{A}^j = \sum_k \lambda^j \boldsymbol{E}_k$, where the $\boldsymbol{E}_k = (\boldsymbol{F}_k + \widetilde{\boldsymbol{F}}_k)/2$ are real. The reader can verify by following these steps that the elements of $\boldsymbol{E}_k$ are as given by (26). QED

Insert now (24) in (23), and note that $\boldsymbol{DU} = \mathbf{0}$ because all the terms of $\boldsymbol{U}$ are identical. The result can be written as:

$$\boldsymbol{\xi_i} = \boldsymbol{D\epsilon_i} = \sum_{j=1}^{i-1}\sum_{k=2}^{N} \lambda_k^j \boldsymbol{DE_k\nu_j} + \sum_{k=2}^{N} \lambda_k^i \boldsymbol{DE_k\epsilon_0}. \tag{27}$$

The last term of this expression vanishes if the initial errors are zero. Now, to simplify the notation, define $\lambda \doteq \max_{k>1}(|\lambda_k|)$. Note that $\lambda < 1$ if $\alpha\Delta t \in (0, 0.5)$, as per Lemma 1; that (27) is a linear combination of the (u.i.v.) noise terms; and that the eigenvalues and the (matrix) weights only depend on $\alpha\Delta t$, and N. These observations are now used in the proof of the following result.

THEOREM: *If*: (*i*) $\alpha\Delta t \in (0, 0.5)$, (*ii*) *the noise terms are uncorrelated with identical variance,* σ^2*, and* (*iii*) *the initial errors are zero, then the quantity* $\mathrm{var}(s_{n,i}) \equiv \mathrm{var}(\xi_{n,i})$ *is bounded by a constant that only depends on* N, $\alpha\Delta t$ and σ^2. Furthermore,

$$\mathbf{cov}(\boldsymbol{\xi_i}) = \sigma^2 \sum_{k=2}^{N} \left(\frac{\lambda_k^2 - \lambda_k^{2i}}{1-\lambda_k^2}\right)(\boldsymbol{DE_k})(\boldsymbol{DE_k})^T, \tag{28}$$

where the eigenvalues and the weights on the RHS are given by Lemma 2.

Proof: The terms of (27) are uncorrelated. Therefore the covariance matrix of their sum is the sum of the individual covariance matrices. So, take covariances on both sides of (27) with $\boldsymbol{\epsilon_0} = \mathbf{0}$, and the result is

$$\mathbf{cov}(\boldsymbol{\xi_i}) = \sum_{j=1}^{i-1}\sum_{k=2}^{N} \lambda_k^{2j}\sigma^2 (\boldsymbol{DE_k})(\boldsymbol{DE_k})^T, \tag{29}$$

This reduces to (28) because $\lambda_k^2 < 1$ by virtue of Lemma 1.

Now, to see that the variances are bounded, use $d_{k,n}$ for the n^{th} diagonal term of $(\boldsymbol{DE_k})(\boldsymbol{DE_k})^T$ and write:

$$\mathrm{var}(\xi_{i,n}) = \sum_{j=1}^{i-1}\sum_{k=2}^{N} \lambda_k^{2j}\sigma^2 d_{k,n}. \tag{30}$$

Note, the three factors on the right side of (30) are non negative. Thus, if we define $d \doteq \max_k(d_{k,n})$, and recall that $\lambda_k^{2j} \le \lambda^{2j} < 1$, we can write:

$$var(\xi_{i,n}) \le \sum_{j=1}^{i-1}\sum_{k=2}^{N} \lambda^{2j}\sigma^2 d \le \frac{\lambda^2}{1-\lambda^2}(N-1)\sigma^2 d < \infty. \tag{31}$$

Since λ and d are only functions of N and $\alpha\Delta t$, this concludes the proof. QED

COROLLARY: *If the noise errors are bounded by a quantity* $\bar{\nu}$*,then the deviations from the spacings are also bounded.*

Proof: Apply the L_1 norm to both sides of the (27) with $\boldsymbol{\epsilon_0} = \mathbf{0}$. Then use the triangle inequality to derive the upper bound by applying the norm to the individual terms of the right hand side. The result is:

$$\|\boldsymbol{\xi_i}\| \leq \sum_{j=1}^{i-1} \sum_{k=2}^{N} \left|\lambda_k^j\right| \|\boldsymbol{DE_k \nu_j}\| \, . \tag{32}$$

If m_k is the largest absolute value of all the elements of $\boldsymbol{DE_k}$, then $\|\boldsymbol{DE_k \nu_j}\| \leq Nm_k\bar{\nu}$. Thus,

$$\|\boldsymbol{\xi_i}\| \leq \sum_{j=1}^{i-1} \sum_{k=2}^{N} \lambda^j N m_k \bar{\nu} \leq \frac{\lambda}{1-\lambda} N\bar{\nu} \sum_{k=2}^{N} m_k < \infty. \quad \text{QED} \tag{33}$$

The corollary is useful because it holds for any joint distribution of the error terms; even if it they are time- and bus-dependent.

Appendix B. Results for forward-looking control

The steps of Section 2.3 are repeated with the following forward-looking law:

$$v'_n \doteq \bar{v} + O_n(\boldsymbol{s}, S), \qquad n = 1, 2, ...N, \tag{34}$$

where $\boldsymbol{s}$ the vector of spacings and $O_n(\boldsymbol{s}, S)$ is:

$$O_n(\boldsymbol{s}, S) \doteq [\lambda b\bar{v}(s_n - S) + \alpha s_n - \delta]/(1 - \lambda b s_n), \qquad n = 1, 2, ...N, \tag{35}$$

The law is forward-looking because it only includes the forward spacing as a state variable. Similar manipulations as in Section 2.3 yield the dynamic commercial speed relation:

$$c'_n \approx \bar{c} + \alpha s_n - \delta, \tag{36}$$

and the quasi-equilibrium trajectories:

$$y_n :\approx y_n \oplus \bar{c}'\Delta t \oplus \alpha s_n \Delta t \oplus \nu_n, \qquad n = 1, 2, ...N. \tag{37}$$

As before, now express (37) in terms of deviations, ϵ_n, from the equlibrium trajectories (10), still recognizing that $s_n = S + \epsilon_{n\ominus 1} - \epsilon_n$, to obtain:

$$\epsilon_n :\approx \epsilon_n \oplus (\alpha\Delta t)\epsilon_{n\ominus 1} \oplus (1 - \alpha\Delta t)\epsilon_n \oplus \nu_n, \qquad n = 1, 2, ...N. \tag{38}$$

For this system of equations, the variance of the coefficients of the dynamic equation is $(\alpha\Delta t)(1 - \alpha\Delta t) \approx \alpha\Delta t$ if Δt is small. Therefore, as in Sec. 3.1, the variance of the spacings can be approximated by:

$$var(s_n) \approx (\alpha\Delta t)var(\nu_n) = r^2/\alpha, \qquad n = 1, 2, ...N. \tag{39}$$

It is still reasonable to choose δ to be about three standard deviations of G, which in the present case is: $G \doteq \lambda b\bar{v}(s_n - S) + \alpha s_n$. Since $var(G) = [\lambda b\bar{v}\alpha + (\lambda b\bar{v})]^2 var(s_n) \approx [\lambda b\bar{v}\alpha + (\lambda b\bar{v})]^2 r^2/\alpha$, we choose:

$$\delta \approx 3[\alpha + \lambda b\bar{v}]r/\alpha^{1/2}. \tag{40}$$

The value of α that minimizes this expression is

$$\alpha^* \approx \lambda b \bar{v}, \tag{41}$$

The optimum reduction in average commercial speed is therefore:

$$\delta^* \approx 6.0\, r (\lambda b \bar{v})^{1/2}, \tag{42}$$

and the variance of the spacing that arises under this type of control is:

$$var(s_n^*) \approx r^2/(\lambda b \bar{v}). \tag{43}$$

Note how (42) and (43) are inferior to the results obtained with two-way looking.

References

Barnett, A., 1974. On controlling randomness in transit operations. Transportation Science 8(2), 102-116.

Daganzo, C.F., 1997. Schedule instability and control. In: Daganzo, C.F., Fundamentals of Transportation and Transportation Operations, Elsevier, New York, N.Y. pp. 304-309.

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

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

Hickman, M.D. 2001 An analytic stochastic model for the transit vehicle holding problem. Transportation Science 35(3), 215-237.

Ignall, E., Kolesar, P., 1974. Optimal dispatching of an infinite capacity shuttle: Control at a single terminal. Operations Research 22(5), 1008-1024.

Newell, G.F., 1974. Control of pairing of vehicles on a public transportation route, two vehicles, one control point. Transportation Science 8(3), 248-264.

Newell, G.F., 1977. Unstable Brownian motion of a bus trip. Statistical Mechanics and Statistical Methods in Theory and Applications (ed. U. Landman), Plenum Press, 645-667.

Newell, G.F. and Potts, R.B., 1964. Maintaining a bus schedule. Proc. 2nd Australian Road Research Board, Vol 2, pp. 388-393.

Osuna, E.E. and Newell, G.F., 1972. Control strategies for an idealized bus system. Transportation Science 6(1), 52-71.

Pilachowski, J., 2009. An approach to reducing bus bunching. PhD thesis Department if Civil and Environmental Engineering, Univ. of California, Berkeley, California.

Zhao, J., Dessouky, M. and Bukkapatnam, S., 2006. Optimal slack time for schedule-based transit operations. Transportation Science 40(4), 529-539.