

Research Article

How Well Does the Traffic System Protect Transit from Congestion? Measuring Route-Level Costs That Congestion Imposes on Transit Operators and Users

Peter G. Furth ¹ and Ahmed T. M. Halawani^{1,2}

¹Northeastern University, Boston, MA, USA

²Taibah University, Medina, Saudi Arabia

Correspondence should be addressed to Peter G. Furth; pfurth@coe.neu.edu

Received 22 December 2017; Revised 28 May 2018; Accepted 3 June 2018; Published 28 June 2018

Academic Editor: Martin Trépanier

Copyright © 2018 Peter G. Furth and Ahmed T. M. Halawani. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As transit agencies and road owners adopt the objective of protecting transit from congestion, it becomes important to have a method for measuring the cost that congestion imposes on transit. Congestion impacts transit both by lowering average speed and by increasing service unreliability. Altogether, five congestion impacts were identified: increased running time and recovery time for transit operators and increased riding time, waiting time, and buffer time for passengers. A methodology for estimating those impacts was developed using automatic vehicle location data. The basic approach was to compare the impact variables during various periods of the week against a base period when there is no congestion (late night and early morning), making adjustments to account for differences in demand that affect running time apart from congestion. The methodology was successfully applied to a sample of 10 bus routes in the Boston area. The cost of congestion on the sample routes was found to range from \$1 to \$2 per passenger, with annual costs as great as \$8 M per year on some routes. Of the total congestion cost, just under 20% applies to the operator, with the remainder applying to passengers. And while the operator is mainly affected by increased average delay, passengers are mainly affected by worsening service reliability.

1. Introduction

Measuring transit performance is more and more becoming accepted as a necessary part of improving service quality [1], with many transit agencies publishing “report cards” to guide improvements and increase accountability (for example, [2]). A parallel question that deserves to be asked is the following: how well does the traffic system protect transit from congestion? On many routes, traffic congestion seriously degrades the quality of transit service, lowering operating speed and worsening service unreliability. This is not news, of course; however, if the congestion impact is not measured, there is little incentive or guidance for improvement; the situation can be allowed to quietly worsen and worsen.

Brussels’ transit operator STIB sets an example in this regard. Their contract with the regional government includes bonuses and penalties tied to route performance [3]. Finding

that many routes could not achieve service reliability targets due to traffic congestion, they began a program to identify “black spots” where congestion was seriously hindering buses and trams. Each year they issue a report on the 50 worst black spots, urging the regional authority to make improvements such as bus lanes [4].

Most transit agencies, however, do not measure or report on traffic congestion because it is exogenous and beyond their control and therefore not part of *their* performance. However, there are treatments that the public agencies that own and control the streets and traffic signals—hereafter the road owner—can apply to protect transit from congestion, such as retiming traffic signals, traffic diversion, bus lanes and queue jump lanes, parking and turn restrictions, and signal priority. In much of Europe, protecting transit from congestion and giving transit priority at traffic signals are a well-accepted responsibility of road owners [5, 6]. And, in an

increasing number of cities in North America, giving priority to transit is an explicit policy [7], and in many others it is recognized as a worthy goal.

Where there is a shared desire on the part of the transit agency and the road owner to improve the quality of public transportation, they need methods to systematically evaluate how much traffic congestion is affecting transit service. Quantifying impacts is needed to identify the problem, to justify investments that give transit priority, and to evaluate projects. The objective of this study was to test for one transit agency, the Massachusetts Bay Transit Authority (MBTA), the feasibility of using readily available automatic vehicle location (AVL) data to calculate the costs that traffic congestion imposes on transit routes, considering impacts to both passengers and the operator, and considering impacts stemming from both increased running time and increased running time variability.

2. Background

Manual measurements made by a person riding the bus have long been used to divide the route cycle into time spent at stops and time spent between stops [8]. However, this process is too labor intensive to be routinely used. Automated data collection systems, mainly AVL and automatic passenger counters (APC), offer new possibilities for measuring transit performance.

With AVL, the main type of data used for performance measurement is records of arrival and departure time at timepoints, which are typically 6 to 10 minutes' travel time apart. Arrival and departure times are typically based on entering and leaving a zone defined by a 30 m radius around the timepoint. Vehicle location typically uses GPS, sometimes supplemented with dead reckoning for travel in zones with unreliable GPS signals such as between tall buildings, in tunnels, and under roofs [9]. APC data is at the stop level and includes records of when doors open and close; however, often only a small fraction of the fleet is instrumented.

The large sample size afforded by automated data makes it possible to measure variability as well as mean of running time and headway. Traffic congestion affects not only average speed; it also affects travel time and headway variability. Litman has argued that service reliability is important to passengers, but because it is harder to measure, it tends to be undervalued, not attracting the investment it should [10].

Many researchers have used AVL to analyze bus travel time and travel time variability. Mazloumi, Currie, and Rose [11] used AVL to measure route-level running time variability and to estimate models relating that variability to factors such as the number of intersections. Al-Geneidy, Horning, and Krizek [12] used timepoint data to measure running time variability at the timepoint segment level and headway variability at each timepoint. Stoll, Glick, and Figliozzi [13] used stop-level data from APCs to measure travel time mean and variance at the stop-to-stop segment level and to thus identify hot spots. Rosenblum et al. [14] used AVL data to identify high traffic delay segments in Cambridge, MA, by

comparing running time between two stops with a minimal-traffic-interference running time found by examining running time early in the morning and late at night. They created a composite score for each segment based on excess running time, excess riding time (equal to excess running time multiplied by passenger load), and the difference in cumulative running time standard deviation between the two stops.

Levinson [8] and McKnight et al. [15] have reasoned that congestion delays felt by buses should mirror those felt by other traffic and so have developed models for bus travel time as a function of the travel time for general traffic along with transit-specific data (e.g., passenger boardings, and alightings). McKnight et al. used their model to determine how much bus running time would decrease if traffic went at free-flow speeds.

Three main shortcomings of these past efforts inspire the current research. First, running time analyses using timepoint data includes time spent serving stops between timepoints, which cannot be attributed to traffic congestion. And, at timepoints that are bus stops, while arrival and departure time are recorded, one cannot identify the time spent at the stop versus time waiting in a traffic queue. This is a generic weakness of AVL data as typically configured for US transit systems. It need not be so; we have long advocated for AVL systems to write a record each time the door opens and closes, as does the AVL system used in Eindhoven (Netherlands) [16].

Second, the measures most often used for service unreliability—such as standard deviation of running time, percent of departures that are on time, or an unreliability index—are not measures that can be directly converted into passenger or operator costs. Furth and Muller [17] have shown that service reliability impacts on passengers can be monetized in terms of additional waiting time and buffer travel time (also called potential waiting time) and that the operating cost impact of service unreliability can be monetized as the extra recovery time operators have to build into their schedules.

Third, none of these studies has attempted to put a cost on the impact of traffic congestion. We know of one transit operator, STIB (Brussels), which estimates the cost of congestion by simply comparing scheduled running time across the day with scheduled running time during a low-traffic period such as Sunday morning. The justification for using scheduled running time is that schedules are regularly updated to reflect actual running times. This is an assumption that can be tested by comparing with a congestion cost estimate drawn from actual performance data. Apart from this, a drawback to their method of comparing against a low-traffic period is that passenger demand per vehicle in such periods is typically lower than average, so that some of the observed difference in running time is due to more time spent at stops and not to traffic congestion. (Adjusting performance for differences in demand is one of the challenges this research aims to address.) In addition, STIB's practice does not attempt to estimate how congestion increases service unreliability, with its attendant costs.

Interval	Day & Period			Interval
	Weekday	Saturday	Sunday	
10:00 PM – 6:29 AM	0 Low_Traffic			10:00 PM – 7:59 AM
6:30 AM – 6:59 AM	5 Shoulder/Eve.			
7:00 AM – 8:59 AM	1 AM Peak	6 SAT_Morning	7 SUN_Morning	8:00 AM – 11:59 AM
9:00 AM – 1:29 PM	2 Midday Base	8 Afternoon_WKEND		12:00 PM- 5:59 PM
1:30 PM – 3:59 PM	3 Midday School			
4:00 PM – 6:29 PM	4 PM Peak	9 Evening_WKEND		6:00 PM – 9:59 PM
6:30 AM – 9:59 PM	5 Shoulder/Eve.			

FIGURE 1: Periods of the week.

2.1. Data Description. The MBTA's AVL system creates three main types of records: heartbeat, timepoint, and stop announcement records. Heartbeat records give bus location every 60 s, but because they are "location-at-time" data, they are undesirable for operations analysis because recorded locations vary from trip to trip, making it impossible to aggregate or compare data without introducing approximation errors.

Timepoint records, which are intended for performance analysis, indicate entry and exit time (called arrival and departure time) for a 30-m radius zone centered at timepoints, which are always bus stops. They include many useful identifiers, including route, route variation, direction, vehicle, and operator ID. They also indicate the scheduled time (arrival time if the timepoint is the last stop, departure time elsewhere), the scheduled headway, and the actual variation from scheduled time and headway, eliminating the need for later matching.

However, timepoint records are unsuitable for running time analysis because data at terminal stops is unreliable, and on MBTA routes the first and last timepoint segments can represent 25 percent of the route or more. Reasons for data unreliability include buses stopping multiple places within the terminal area, opening and closing doors several times, and roof structures causing GPS occlusion [9].

Stop announcement records, intended only as a verification tool, turn out to be the key to measuring running time. A record is made of every announcement. Fields in the announcement record include the announcement type, identifiers for the route, vehicle, trip, and driver, a time stamp, GPS location, and a stop sequence number. Of interest to this research are external announcements, which are made at every stop when the front door first opens and are repeated every 30 s while the front door remains open. Estimated running time, excluding dwell at the terminals, can then be estimated as the difference in time from the last record written at the origin terminal to the first record at the destination terminal, plus 15 s. The 15 s correction is needed because the door closing time at the first stop could have been up to 30 s later than the last announcement there; with the 15 s

correction, the estimate is unbiased and the approximation error is uniformly distributed on the interval ± 15 s.

Stop announcement data also provide a count of how many stops were made on each trip, since external announcements are only made when the doors open.

Route- and stop-level demand information can be obtained from automatic passenger count (APC) data, which is collected on 12% of the fleet. Because the demand data needed in this analysis can be aggregated over a year, this sampling rate is more than adequate.

3. General Approach to Identifying Congestion Impacts

Following the example of Brussels and Cambridge described earlier, our approach to identifying congestion impacts is to compare different periods of the week against a base period during which there is little traffic. We then make corrections to account for differences in demand between each period and the base period.

Figure 1 indicates the periods of the week used for this analysis. Trips are classified based on their start time. Period 0, the base period, consists of early morning and late night service on both weekday and weekend.

4. Estimating Congestion Impacts

The components of congestion impact and formulas used to measure them are summarized in Table 1. The variables and formulas used in Table 1 are explained in the corresponding subsections that follow.

4.1. Running Time Impact. For each period, adjusted running time is calculated by eliminating running time elements related to passenger demand:

$$\text{AdjRT}_p = \text{RT}_p - \gamma(N\text{Stop}_p - 1) - \beta\text{Ons}_p - \alpha\text{Offs}_p \quad (1)$$

TABLE 1: Congestion impact components.

Component	Unit	Unit cost (\$/hr)	Congestion Impact (min)	Congestion impact details
Running time (RT)	Bus trip	\$108	$\text{AdjRT}_p - \text{AdjRT}_o$	Equation (1)
Riding time or in-vehicle time (IVT)	Passenger	\$12	$\text{AdjIVT}_p - \text{AdjIVT}_o$	$\text{AdjIVT}_p = 0.4\text{AdjRT}_p$
Recovery time	Bus trip	\$108	$\text{Recov}_p - \text{IdealRecov}_p$	Recov_p : equations (2)-(7) IdealRecov_p : equation (8)
Waiting time	Passenger	\$18	$\text{ExcessWait}_p - \text{IdealExcessWait}_p$	<i>For short headway service:</i> ExcessWait_p : equation (9) IdealExcessWait_p : equation (16) <i>For long headway service:</i> ExcessWait_p : equation (17) $\text{IdealExcessWait}_p = \text{ExcessWait}_o$
Buffer time	Passenger	\$9	$\text{BufferTime}_p - \text{IdealBufferTime}_p$	$\text{BufferTime}_p = 0.75 \text{Recov}_p$ $\text{IdealBufferTime}_p = 0.75 \text{IdealRecov}_p$

where

AdjRT_p is adjusted running time for period p (min).

RT_p is mean running time for period p (AVL data) (min).

NStop_p is mean number of stops per trip in period p (AVL data).

Ons_p is mean boardings per trip in period p , excluding terminals (APC data).

Offs_p is mean alightings per trip in period p , excluding terminals (APC data).

α , β , and γ are coefficients for alightings, boardings, and number of stops, respectively.

A dwell time model was estimated to determine boarding and alighting time coefficients. There were 14,289 stop records from the MBTA's APC data for the months of September, October, and November, 2012, each containing the time that the doors were open (the dependent variable), ons, and offs. Estimates are $\beta = 0.0725$ min (4.35 s) and $\alpha = 0.0312$ min (1.87 s), close to other estimates reported in the literature. R^2 was 0.64.

The coefficient $\gamma = 0.235$ min (14.1 s) was determined analytically. Lost time for deceleration and acceleration was assumed to be 3.5 s and 7.0 s, respectively, based on uniform deceleration and acceleration at rates of 3.0 mph/s and 1.5 mph/s, respectively, and a final speed of 21 mph that represents mix of buses topping out at 30 mph and buses whose speed is constrained by impending stop lights or turns. (Note that, under constant acceleration to or from a stop, acceleration delay is half the acceleration time.) The final value for lost time at a stop, 14.1 s, is very close to the 15 s given in the *Transit Capacity and Quality of Service Manual* [18]. Lost time with the door fully open was taken to be 1.6 s, the constant estimated in the dwell time model; an additional 2 s of lost time was also assumed to account for time between when doors open/close and wheels stop/start rolling.

In (1), we subtract 1 from NStop_p because the AVL data yields measurements of running time from when the doors close at the first stop to when they open at the last stop.

The difference in adjusted running time between any given period p and the base period is then attributed to traffic congestion. The unit cost for running time used was \$108/h, which is the MBTA's reported bus operating cost per vehicle hour for 2012 multiplied by 0.70, which is roughly the fraction of operating cost that is appropriately allocated to vehicle-hours.

4.2. Riding Time Impact. Passenger in-vehicle time (IVT) is the same as running time, except that it extends over only part of the route, and should be adjusted the same way to subtract out the demand-related effects of stopping and serving passengers. The difference between adjusted travel time in any period and the base period can be attributed to traffic congestion.

In principle, adjusted IVT can be calculated by segment, weighted by passenger load, and then aggregated. A simpler, approximate way to estimate adjusted IVT is to use a ratio of average passenger trip length to vehicle trip length. The congestion-induced increase in passenger IVT should be roughly the same as the increase in vehicle running time multiplied by this ratio. For this study, we used a default ratio of 0.4 based on a study by Furth [19]. With further analysis of APC data, it may be possible to calculate an appropriate ratio for each route.

4.3. Recovery Time Impact. Traffic congestion not only lowers average speed; it also tends to increase running time variability. The direct operating cost impact is the increased recovery time transit agencies have to build into their schedules. The impact for period p , then, is the difference ($\text{Recov}_p - \text{IdealRecov}_p$) where Recov_p is needed recovery time in period p and IdealRecov_p is recovery time that would be needed in period p if there were no congestion.

The MBTA's policy is to provide enough recovery time in any given period so that the sum of scheduled running time

plus recovery time equals the 95th-percentile running time, and because 1.64 is the coefficient associated with the 95th percentile in a normal approximation,

$$\text{Recov}_p = 1.64\text{sqrt}(V\text{FromSch}_p) \quad (2)$$

where $V\text{FromSch}_p$ is mean squared schedule deviation, informally called the variance from schedule, given by

$$V\text{FromSch}_p = \frac{1}{n-1} \sum (t_i - s_i)^2 \quad (3)$$

where t_i is observed running time on trip i , s_i is scheduled running time on trip i , and the sum is over all the trips in a period (over many days). If AVL announcement records were matched to scheduled trips, $V\text{FromSch}_p$ could be calculated directly. However, they are not, and so $V\text{FromSch}_p$ must be estimated.

Two related quantities that *can* be calculated for any given period are the within-period running time variance $V_p(\text{RT})$, given by

$$V_p(\text{RT}) = \frac{1}{n-1} \sum_i \sum_j (t_{ij} - \bar{t}_p)^2, \quad (4)$$

and the within-period scheduled running time variance $V_p(S)$ (relevant because scheduled running time can vary within a period p) given by

$$V_p(S) = \frac{1}{n-1} \sum_i (s_i - \bar{s}_p)^2 \quad (5)$$

where i is the trip index and the sum over i covers all the trips within period p , j is a day index and the sum over j covers all days in the analysis horizon, t_{ij} is running time on trip i for day j , \bar{t}_p is the grand mean running time for the period,

s_i is the scheduled running time for trip i , and \bar{s}_p is the mean scheduled running time for the period.

Beginning with (3), we can write

$$\begin{aligned} V\text{FromSch}_p &= E(t_i - s_i)^2 \\ &= E\left\{\left[\left(t_i - \bar{t}_p\right) - \left(s_i - \bar{s}_p\right) + \left(\bar{t}_p - \bar{s}_p\right)\right]^2\right\} \\ &= E\left(t_i - \bar{t}_p\right)^2 + E\left(s_i - \bar{s}_p\right)^2 \\ &\quad - 2E\left[\left(t_i - \bar{t}_p\right)\left(s_i - \bar{s}_p\right)\right]. \end{aligned} \quad (6)$$

(Expanding the square yields two other cross-product terms, but their values are zero.) Rewriting the last term in terms of r_{ts} , the correlation between scheduled and actual running time in the period of analysis, variation from scheduled running time is given by

$$\begin{aligned} V\text{FromSch}(RT)_p &= V_p(\text{RT}) + V_p(S) + \left(\bar{s}_p - \bar{t}_p\right)^2 \\ &\quad - 2r_{ts}\sqrt{V_p(\text{RT})}\sqrt{V_p(S)}. \end{aligned} \quad (7)$$

Because r_{ts} cannot be measured with unmatched data, it is both conservative (i.e., to avoid overestimation) and reasonable to assume strong correlation, since the aim of the scheduling function is to schedule longer running times when actual running times are longer. For the following case study, we assume $r_{ts} = 0.8$. With this approximation, (3)–(7) can be used to calculate needed recovery time.

To find the IdealRecov_p , we reason that if there were no traffic congestion in period p , its variance from schedule would be the same as it is period 0, plus variance that arises due to increased passenger arrivals and increased stops. Assuming Poisson passenger arrivals, for which the variance in number of passengers equals the mean, and using direct measurements of $V(\text{NStop})$ using AVL data,

$$\text{IdealRecov}_p = 1.64\sqrt{V\text{FromSch}_o + (\alpha + \beta)^2 (\text{Ons}_p - \text{Ons}_o) + \gamma^2 [V_p(\text{NStop}) - V_o(\text{NStop})]}. \quad (8)$$

4.4. Increased Passenger Costs due to Unreliability. Just as service unreliability forces operating agencies to schedule recovery time, it forces passengers to schedule additional travel time, called buffer travel time; it also increases their average waiting time. Passenger waiting time depends on whether headways are short (for this study, under 13 minutes) or long; in the former case, passengers are assumed to arrive without regard to a schedule, while, in the latter case, they are expected to aim for a particular scheduled departure.

4.5. Waiting Time Impact with Short Headway Service. With short headway service, it is well known that

$$\text{excess waiting time} = \frac{V(h)}{2H} \quad (9)$$

where h is headway and H is mean headway. For this application, the headway of interest is at the boarding stop, which of course varies along the route, but for which a representative stop can be chosen. Among AVL record types, timepoint records are most appropriate for measuring headways because they are matched to a scheduled trip. Because boardings tend to take place near the beginning of a route, the representative boarding stop was the timepoint (excluding the start point) closest to being the quarter-way point along the route.

Headway variance has three main sources: operational control (e.g., departing on time, maintaining consistent deceleration and acceleration rates), randomness in passenger arrivals, and randomness in traffic congestion. The base period's headway variance can be assumed to contain the

operational control component and some of the demand component, and so the headway variance that would prevail in period p if there were no congestion is the base period's variance plus the variance that would stem from the increase

$$t_b = \text{running time to the boarding stop } b \text{ within the period of interest} \quad (10)$$

and let $t_{b,1}$ and $t_{b,2}$ be the realization of t_b for a pair of successive trips. Then the headway at b will be

$$h = h_{\text{dispatch}} + (t_{b,2} - t_{b,1}) \quad (11)$$

where h_{dispatch} is dispatch headway (itself a random variable). Treating running time on successive trips as independent, the variance in headway change arising from running time variance is

$$V(h) = V(h_{\text{dispatch}}) + 2V(t_b). \quad (12)$$

We reason that, on the segment between the dispatch terminal and the representative boarding point, roughly 25% of stops, 50% of boardings, and 15% of alightings will have been made. With those assumptions, running time to point b during period p can be expressed as

$$t_{b,p} = t'_{b,p} + 0.25\gamma N\text{Stop}_p + 0.5\beta \text{Ons}_p + 0.15\alpha \text{Ons}_p \quad (13)$$

where $t'_{b,p}$ is running time component away from stops. If there is no congestion, this running time component will be generic to all periods including the base period, and so the subscript p can be dropped. Calculating the variance of t_b for both period p and period 0, using the Poisson assumption

in demand per vehicle between the base period and the period of interest.

To model the headway variance impact of demand, let

that $V(\text{Ons}) = \text{Ons}$, and rearranging, the variance of t_b if there were no congestion effect is

$$\begin{aligned} V_p(t_b) &= V(t'_b) \\ &+ (0.25\gamma)^2 [V_p(N\text{Stops}) - V_o(N\text{Stops})] \\ &+ (0.5\beta + 0.15\alpha)^2 (\text{Ons}_p - \text{Ons}_o). \end{aligned} \quad (14)$$

Substituting then into (12), headway variances if there were no congestion effect for period 0 and for period p are

$$\begin{aligned} V_{o,\text{Ideal}}(h) &= V(h_{\text{dispatch}}) + 2V(t'_b) \\ V_{p,\text{Ideal}}(h) &= V_{o,\text{Ideal}}(h) \\ &+ (0.25\gamma)^2 [V_p(N\text{Stop}) - V_o(N\text{Stop})] \\ &+ (0.5\beta + 0.15\alpha)^2 (\text{Ons}_p - \text{Ons}_o). \end{aligned} \quad (15)$$

During the base period, there is no short headway service from which $V_{o,\text{Ideal}}(h)$ could be directly measured. A reasonable assumption is that, in the absence of traffic congestion, $V_{o,\text{Ideal}}(h) = 1 \text{ min}^2$, in which case ideal excess waiting time becomes

$$\text{IdealExcessWait}_p = \frac{\{1 + 2(0.25\gamma)^2 [V_p(N\text{Stop}) - V_o(N\text{Stop})] + 2(0.5\beta + 0.15\alpha)^2 (\text{Ons}_p - \text{Ons}_o)\}}{2H} \quad (16)$$

As indicated in Table 1, the difference between excess waiting time and ideal excess waiting time is attributed to traffic congestion. The unit cost, \$18/passenger-h, is set to be 1.5 times the value of in-vehicle time. While it is customary in transportation planning to apply a multiplier of 2 to 2.5 to waiting time, Furth and Muller [17] show that when the indirect waiting costs due to unreliability are accounted for, applying a multiplier of 1.5 for the value of direct waiting time is consistent with standard practice of using a 2.5 multiplier and neglecting the travel time impacts of unreliability.

4.6. Waiting Time Impact with Long Headway Service. With long headway service, passengers are expected to arrive targeting a scheduled departure, and so waiting time depends on the departure time deviation from the schedule. Following [17], we assume that passengers limit their chance of missing

the bus by arriving no later than the 2nd-percentile departure time, so that excess waiting time under long headway service becomes

$$\text{ExcessWait} = \text{DepDev}_{\text{mean}} - \text{DepDev}_{0.02} \quad (17)$$

where $\text{DepDev}_{\text{mean}}$ and $\text{DepDev}_{0.02}$ are the mean and 2nd percentile departure deviation at the boarding stop, respectively.

Because departure deviations can only be calculated at timepoints, the timepoint closest to the quarter point of the route is used as the representative boarding stop.

Because excess waiting time applies only to periods with long headways in which demand is low, no adjustment is made for demand. The full difference between excess wait in period and excess wait in the base period is then attributed to congestion, as indicated in Table 1.

4.7. Buffer Time Impact. Buffer travel time is time beyond the mean travel time that people have to budget for travel due to unreliability. As in [17], we assume that people budget based on the 95th-percentile arrival time at their destination stop—that way, they will be late 5 percent of the time. Buffer travel time is then the difference between the 95th-percentile arrival time and the mean arrival time at the destination stop. This quantity can readily be calculated from timepoint data; however, it is not amenable to adjustment to account for demand, and therefore it is difficult to isolate what part of it can be attributed to traffic congestion.

An approximate way to estimate potential travel time is to take the difference between the 95th-percentile arrival time and the scheduled arrival time at the destination stop. (This is a reasonable approximation for the MBTA, where schedules are based on mean running time.) If the destination stop were the terminal, this would be exactly the same as the needed recovery time. On average, a passenger's destination stop will be toward the end of a route but not at the very end; therefore we will assume that buffer time is 75% of needed recovery time and that ideal buffer time (the buffer time that would be needed if there were no congestion impact) is 75% of the ideal recovery time, as indicated in Table 1.

Consistent with [17], buffer travel time (there called potential waiting time) is given a value multiplier of 0.75, resulting in a unit cost of \$9. The rationale is that this is time that people budget for travel but do not actually spend traveling; on average, they arrive earlier than budgeted at their destination and use that “redeemed” time in a more productive or enjoyable way than sitting in a vehicle.

5. Estimating Running Time and Riding Time Impacts from Schedule Data

A far simpler method for estimating congestion impacts on running time and in-vehicle time is simply to compare scheduled running times against those of a base period. This is reasonable if scheduled running times are regularly updated to reflect actual running times (as they are at the MBTA). The main drawback of this approach is that it offers no way to estimate the impacts of congestion on reliability; this approach can only estimate impacts on the operator and on passengers from a lower average speed.

With this approach, running time data are replaced with schedule data. Passenger count data are still needed as before. Where AVL data were used to count the number of stops, a model-based estimate can be substituted. Let m_p = number of movements (ons and offs) per stop during period p , which can be calculated as

$$m_p = \frac{2\text{Ons}_p}{N} \quad (18)$$

where N is number of stops on the route. Then, using a Poisson model, the probability of a bus stopping at a given stop equals the probability of one or more people wanting to get on or off:

$$P(\text{stop}) = 1 - e^{-m_p}. \quad (19)$$

The expected number of stops per trip in a given period p is then

$$E[N\text{Stop}_p] = N * (1 - e^{-m_p}) \quad (20)$$

which can be used to substitute for $N\text{Stop}_p$.

6. Case Study Results

Congestion-imposed costs were calculated as a case study for ten MBTA routes, of which six are high-frequency “key routes” and two are actually variations of a single route (Route 89). Table 2 shows detailed calculations for selected periods for Route 1. In the column Adjusted RT, the excess over the base period is due to congestion and is substantial for the nonbase periods. Comparing Recov and IdealRecov, the difference (due to congestion) is substantial for the latter two periods. The base and midday periods both have long headways, and the midday period's difference between excess wait ideal excess wait is very large.

Table 3 shows the five congestion impacts on Route 1 by period and direction, aggregated to yield an annual impact. The annual cost of traffic congestion to this route is a little over \$7.2 million, of which about \$1.2 million is a direct operating cost due to longer running times and recovery times, and \$6M is a cost to passengers, as congestion forces them to spend more time in the vehicle, waiting and committing to a travel time budget in case the trip is late or slow.

This large annual congestion cost offers some insight into how much it would be worth protecting Route 1 from congestion. With a 4% discount rate, a 50-year stream of costs at \$7.2M/y has a present value of \$155M. That means that if an investment of \$155M could protect Route 1 from all congestion for 50 years, it would be worth it. Building a subway would alleviate all congestion impacts but would probably cost 10 times that amount. The challenge then is to find lower cost solutions that might eliminate enough congestion impacts to make them worthwhile.

Table 4 shows impacts on the 10 routes tested. The overall cost per passenger is remarkably similar over the first seven routes, roughly \$1.80 per passenger, while the last three routes (one urban, one suburban, and one a mix) see congestion costs around \$1.20 per passenger. This finding can be used to help justify using autobased fees (e.g., gasoline taxes, congestion charges) to subsidize transit. Like Route 1, the other routes show substantial impacts in all categories, with passenger costs accounting for the majority of the impact and operating costs accounting for about 18%.

Finally, Figure 2 shows a comparison between cost impacts estimated using AVL data and impacts estimated using schedule data instead. This comparison includes only running time and riding time impacts, because reliability-related impacts cannot be estimated from schedule data. Given this limitation, however, the fit appears to be very good, suggesting that as an interim measure, a transit agency could estimate at least part of the traffic congestion impact it suffers quite easily. The more detailed analysis of this case study suggests that the running time impact accounts for 73% of the total operating cost impact and that the riding time impact

TABLE 2: Elements of impact calculation for Route 1 inbound, selected periods.

<i>Period</i>	<i>Net Ons</i>	<i>mean RT (min)</i>	<i>mean NStop</i>	<i>σ(Nstop) (min)</i>	<i>Adjusted RT (min)</i>	<i>Recov (min)</i>	<i>Ideal Recov (min)</i>	<i>headway (min)</i>	<i>σ(headway) (min)</i>	<i>mean DepDev (min)</i>	<i>2-p DepDev (min)</i>	<i>Excess wait (min)</i>	<i>Ideal excess wait (min)</i>
<i>base</i>	22.1	28.4	16.7	4.9	22.4	6.9	6.9	long	n/a	0.4	-2.5	2.9	2.9
<i>AM peak</i>	64.6	43.0	24.7	3.2	30.9	9.9	6.8	9.5	4.7	n/a	n/a	1.2	0.1
<i>midday</i>	64.9	43.0	24.3	3.4	31.1	11.1	6.8	long	n/a	0.7	-5.4	6.1	2.9
<i>PM peak</i>	57.4	46.0	24.3	3.9	34.9	14.3	6.9	8.2	5.8	n/a	n/a	2.0	0.1

TABLE 3: Congestion impacts to Route 1 by period-direction and overall.

Direction	period	headway type	trips/yr	passengers/trip	Run time impact per trip (min)	Recovery time impact per trip (min)	In-vehicle time impact per passenger (min)	Wait time impact per passenger (min)	Buffer time impact per passenger (min)	Total
in	0	long	7300	26.3	0	0	0	0	0	
in	1	short	3289	71.3	8.4	3.1	3.4	1.1	2.3	
in	2	long	5060	71.2	8.7	4.3	3.5	3.1	3.2	
in	3	short	3289	75.6	15	6.9	6	1.9	5.1	
in	4	short	4554	63.6	12.5	7.4	5	2	5.6	
in	5	short	6072	50.4	4.4	0	1.8	1.4	0	
in	6	short	1176	48.9	5.1	1.4	2	2.1	1.1	
in	7	long	672	50.8	3.6	1.8	1.4	1.3	1.4	
in	8	short	3192	59	9.4	3	3.7	1.2	2.2	
in	9	long	1736	44.8	5.5	0.8	2.2	3.1	0.6	
out	0	long	8171	28.2	0	0	0	0	0	
out	1	short	3036	66.9	4.3	4.7	1.7	1.5	3.5	
out	2	long	5060	73.8	5.6	2.9	2.3	2	2.2	
out	3	short	3542	68.7	8.2	2.4	3.3	2.4	1.8	
out	4	short	4807	70.1	11.9	9.2	4.8	2.2	6.9	
out	5	short	6325	50.8	3.6	2.6	1.4	1.7	2	
out	6	short	1232	40.9	5.6	2.8	2.2	1.8	2.1	
out	7	long	728	53.2	3.9	8	1.6	0.1	6	
out	8	short	3192	58.8	7.9	6.2	3.1	1.4	4.6	
out	9	long	1736	47.2	3	0	1.2	0.5	0	
Annual impact (vehicle-h or passenger-h per year)					7,389	3,817	191,423	113,634	188,023	
unit cost (\$/vehicle-h or \$/passenger-h)					108	108	12	18	9	
Annual Impact (\$/yr)					798,023	413,699	2,297,076	2,045,403	1,692,202	7,246,403

TABLE 4: Congestion impacts on ten routes.

Route	Passengers (1000/y)	running time impact (\$1000/y)	recovery time impact (\$1000/y)	riding time impact (\$1000/y)	wait + buffer impact (\$1000/y)	Total (\$1000/y)	Cost per passenger
1	4,055	801	414	2,297	3,738	7,249	\$ 1.79
23	3,714	913	283	2,107	3,102	6,405	\$ 1.72
28	4,362	789	303	2,387	4,269	7,749	\$ 1.78
39.3	4,024	971	358	2,654	3,383	7,366	\$ 1.83
66.6	4,008	1,051	118	3,427	3,525	8,121	\$ 2.03
77	2,271	660	225	1,058	2,116	4,058	\$ 1.79
89	508	149	66	187	450	853	\$ 1.68
89.2	673	88	112	147	531	877	\$ 1.30
99.7	431	132	56	147	118	453	\$ 1.05
9	1,867	360	270	613	1,112	2,354	\$ 1.26
Total	25,914	5,914	2,205	15,023	22,343	45,484	\$ 1.76

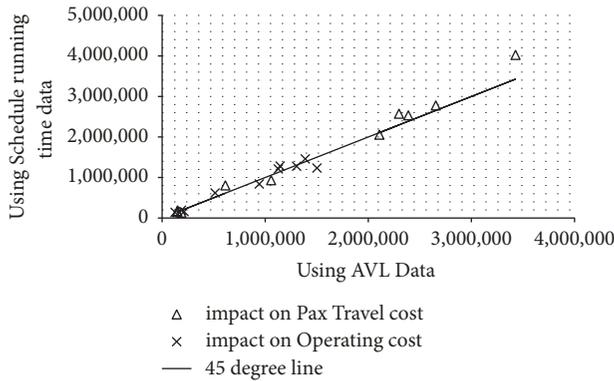


FIGURE 2: Congestion impact estimates made using schedule data versus AVL data.

accounts for 40% of the passenger cost impact; these ratios could be used to expand results obtained by simple schedule analysis.

7. Conclusion and Comments

This research has shown that it is feasible to measure the costs that traffic congestion imposes on transit operators and users on a route-level basis, including costs associated with lowering average travel speed and those associated with increased service unreliability. The method we have developed is one that could be automatically applied annually (for example) to all routes, using as input available AVL, APC, and schedule data.

This methodology successfully addresses the challenges of accounting for the demand effect on running time and of measuring service unreliability impacts in terms of costs to the transit operator and to passengers. Altogether, congestion impacts come in five categories: increased running time and recovery time for transit operators, and increased riding time, waiting time, and buffer time for passengers.

At the same time, the need to make assumptions and modeling approximations makes the outputs of this methodology less robust than if impacts were measured more directly. One improvement would be adding timepoints a short distance from the terminals, where vehicle jockeying and occlusion are not an issue but still close enough to measure running time over almost the whole route. Another would be programming the AVL system to make records when doors open and close, together with identifiers that would allow those records to be matched to timepoint data, so that dwell time could be readily deducted from running time, eliminating the need to model how changing demand levels affect running time.

The results for the ten bus routes tested indicate that the costs of congestion are substantial, ranging from \$1 to \$2 per passenger and with annual costs as great as \$8 M per year for a route. Just under 20% of the total congestion cost applies directly to the operator; the remainder applies to passengers. Also, costs stemming from congestion's impact on service unreliability are slightly greater overall than those

stemming from its impact on average speed, but they affect the parties differently. For the transit operator, 73% of their congestion impact is due to lower average travel speed, while, for passengers, 60% of their congestion impact is due to unreliability.

Giving priority to transit is increasingly becoming recognized as an appropriate goal for traffic management. While nobody questions that traffic congestion hurts transit, there is a saying in business that only what is measured counts. Applying a methodology such as this one to routinely measure and report how much transit is hurt by congestion will be valuable for communicating the extent of the problem and for guiding and evaluating improvements.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] G. J. Fielding, "Transit performance evaluation in the USA," *Transportation Research Part A: Policy and Practice*, vol. 26, no. 6, pp. 483–491, 1992.
- [2] "Massachusetts Bay Transportation Authority, MBTA Performance Dashboard," <http://www.mbtackontrack.com/performance/index.html#/home>.
- [3] J.-C. Liekendael, P. Furth, and T. Muller, "Service quality certification in brussels, Belgium: A quality process with teeth," *Transportation Research Record*, no. 1955, pp. 88–95, 2006.
- [4] B. Guy, *Les Points Noirs du Réseau Stib*, LaLibre.be, 2014.
- [5] P. G. Furth, *Public Transport Priority for Brussels: Lessons from Zurich, Eindhoven, and Dublin*, Report to the Brussels Capital Region, Université Libre de Bruxelles, 2005.
- [6] Nash, B. Andrew, and D. S. Ronald, *Implementation of Zürich's Transit Priority Program*, Mineta Transportation Institute, San Jose State University, 2001.
- [7] "San Francisco Municipal Transportation Agency. Transit First Policy," <https://www.sfmta.com/transit-first-policy.Policy>.
- [8] H. S. Levinson, "Transit travel time performance," *Transportation Research Record*, vol. 915, pp. 1–6, 1983.
- [9] P. G. Furth, B. Hemily, T. H. J. Muller, and J. G. Strathman, "Using archived AVL-APC data to improve transit performance and management," Tech. Rep., Transportation Research Board, 2006, TCRP Report 113.
- [10] T. Litman, "Valuing transit service quality improvements," *Journal of Public Transportation*, vol. 11, no. 2, pp. 43–63, 2008.
- [11] E. Mazloumi, G. Currie, and G. Rose, "Using GPS data to gain insight into public transport travel time variability," *Journal of Transportation Engineering*, vol. 136, no. 7, Article ID 006007QTE, pp. 623–631, 2010.
- [12] A. M. El-Geneidy, J. Horning, and K. J. Krizek, "Analyzing transit service reliability using detailed data from automatic vehicular locator systems," *Journal of Advanced Transportation*, vol. 45, no. 1, pp. 66–79, 2011.
- [13] N. B. Stoll, T. Glick, and M. A. Figliozzi, "Using high-resolution bus GPS data to visualize and identify congestion hot spots in urban arterials," *Transportation Research Record*, vol. 2539, pp. 20–29, 2016.

- [14] J. L. Rosenblum, D. W. Allen, T. L. Bennett, R. K. Warade, and C. M. Stoughton, "Method for assessing bus delay in mixed traffic to identify transit priority improvement locations in Cambridge, Massachusetts," *Transportation Research Record*, vol. 2533, pp. 60–67, 2015.
- [15] C. E. McKnight, H. S. Levinson, K. Ozbay, C. Kamga, and R. E. Paaswell, "Impact of traffic congestion on bus travel time in northern New Jersey," *Transportation Research Record*, no. 1884, pp. 27–35, 2004.
- [16] P. G. Furth, T. H. J. Muller, J. G. Strathman, and B. Hemily, "Designing automated vehicle location systems for archived data analysis," *Transportation Research Record*, no. 1887, pp. 62–70, 2004.
- [17] P. G. Furth and T. H. J. Muller, "Service reliability and hidden waiting time insights from automatic vehicle location data," *Transportation Research Record*, no. 1955, pp. 79–87, 2006.
- [18] Kittelson and Associates et al., *Transit Capacity and Quality of Service Manual*, Transportation Research Board, 2013.
- [19] P. G. Furth, "Sampling and estimation techniques for estimating bus system passenger-miles," *Journal of Transportation and Statistics*, vol. 8, no. 2, pp. 87–100, 2005.



Hindawi

Submit your manuscripts at
www.hindawi.com

