

Transit Signal Priority Along Arterials Using Advanced Detection

Yann Wadjas and Peter G. Furth

This research developed and tested the concept of advanced detection and cycle length adaptation as a strategy for providing priority for transit vehicles. In a departure from control strategies that rely on detection only a few seconds in advance of the stopline, a control algorithm was developed in which transit vehicles are detected two to three cycles in advance of their arrival at an intersection stopline, and phase lengths were then constrained so that the transit-serving phase was green for a 40-s predicted arrival window. Methods were developed for selecting whether to extend or compress phase lengths to shift a green period to cover the arrival window. Adaptive control was combined with actuated control using traffic density and queue length estimation, transit stopline actuation, and peer-to-peer communication for coordination in the peak travel direction. The method was applied by simulation to Boston, Massachusetts' Huntington Avenue corridor, which is served by a light-rail line running partly in mixed traffic and partly in a median reservation. The prediction/adaptation algorithm resulted in 82% of the trains arriving during the green phase. This control strategy resulted in substantial improvements to transit travel time and regularity with negligible impacts on private traffic and pedestrians, and was found to be more effective than simple preemption.

Giving transit priority at traffic signals is a worthy societal objective. However, its realization is complicated by the difficulty, due to variable dwell times at stops, in predicting when transit vehicles will arrive at an intersection. In most implementations, detection occurs no further from the stopline than the closest upstream stop or signalized intersection. For near-side stops, this allows for nearly no advanced detection; for far-side stops (and intersections with no transit stop), advanced detection is limited by the inter-stop running time, which in typical U.S. applications is 20 to 30 s. This short notice limits the degree to which the traffic signal cycle can be adjusted to serve the coming transit vehicle, especially in view of the often long clearance times needed for pedestrian crossings, an important concern on transit arterials. With such short notice, aggressive priority tactics can seriously disrupt other traffic. For example, at a four-phase intersection with aggressive transit priority in the Netherlands, Furth and Muller (1) found that when the frequency of transit priority requests increased to 12 per hour, the capacity loss was so great that traffic queues and delays to the approaches not serving transit became unbearable. No wonder, then, that most transit signal priority applications in the United States are limited to less aggressive tactics such as short green extensions and sometimes short red truncations.

This research tests the feasibility of a different control approach: detecting transit vehicles two to three signal cycles in advance of their arrival at an intersection and beginning then to make signal

cycle adjustments so the transit phase will be green when the transit vehicle arrives. The longer advanced notice allows the controller to adjust the lengths of many phases by small amounts, limiting disruption to other traffic and respecting pedestrian needs. However, the longer notice carries with it a far greater uncertainty of when the transit vehicle will arrive because of randomness in dwell time and delay at intervening intersections and stops. To overcome this uncertainty, the cycle was adjusted so the transit-serving phase will be green for a rather broad (± 20 s) arrival time window. If the longer detection period is sufficient to allow the controller to provide this window by making several small adjustments to phase lengths, this approach should be able to provide a higher level of service to both transit and other traffic than can be provided with short-term detection.

The general approach of advanced detection and cycle adjustment has been advanced by Larry Head of Siemens/Gardner Systems (2), who in recent conferences has reported success in applying this strategy based on both simulation and field implementation in Salt Lake City, Utah. This approach has also been found to be beneficial in a simulation study of light rail in Houston, Texas (3). This simulation-based research, using the authors' implementation of this general approach in a very different setting, confirms the promise of this approach.

HUNTINGTON AVENUE/E BRANCH CORRIDOR

With VISSIM (4) simulation software, Boston, Massachusetts' Huntington Avenue corridor from the subway portal at Opera Place to Heath Street was modeled, a distance of 2.5 km with eight signalized intersections and eight stops in each direction (see Figure 1). The Massachusetts Bay Transportation Authority operates a light-rail line known as the E Branch of the Green Line at grade from Heath Street to the portal at Opera Place, from which point it continues in a subway through downtown Boston. Huntington Avenue has two lanes plus parking in each direction. The train runs in a median reservation except in the outer 1 km of the corridor, where there is no median and train tracks are in the street.

Weekday traffic and transit operations during the morning-peak period in 1995 were modeled. E Branch trains consisting of two articulated cars run every 9 min in each direction. A scenario with double the passenger volume and average train headways of 4.5 min was also tested. Intersections generally have enough traffic capacity. Most have volume-to-capacity ratios on the critical intersection approach (Figure 1) between 0.5 and 0.7. To test the control strategy under more severe conditions, a scenario is included in which traffic volumes are doubled.

Pedestrian volumes are high as the primary land uses are multi-family housing, major hospitals, and institutions of higher education including Harvard University's medical campus and Northeastern

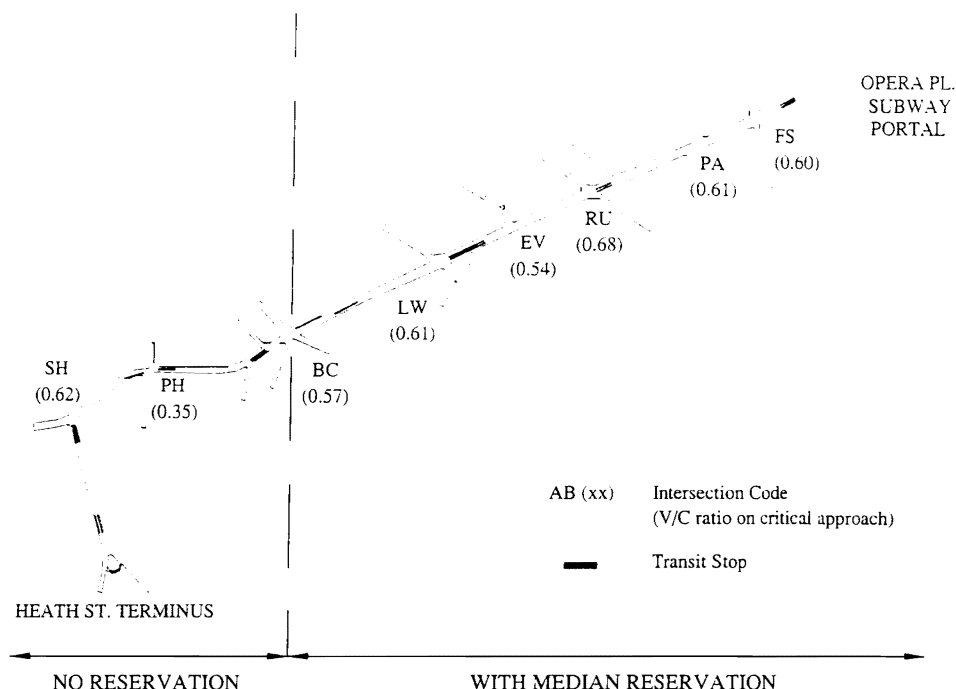


FIGURE 1 Huntington Avenue corridor (V/C = volume-to-capacity).

University. Except at one intersection, pedestrian crossings are concurrent with traffic phases. Because of the city's policy of not relying on the median as a pedestrian refuge, pedestrian clearance time (flashing Don't Walk) across Huntington Avenue is 24 s. Out of concern for pedestrian service, long signal cycles (exceeding 100 s or more) are not permitted.

Current traffic control is a combination of fixed time and semi-actuated with fixed cycle length. Most signals have three phases. Trains do not have their own phase but run concurrently with general traffic. Where Huntington Avenue has a median, many signals have permitted left turns from Huntington Avenue, which conflict with train movements in both directions. Besides posing a safety hazard, these left-turning vehicles often wait in the median for a break in opposing traffic, blocking one or both pairs of tracks. There is no coordination between intersections, with each having its own cycle length. The potential for coordination is weak because of the large turning volumes and the fact that Huntington Avenue serves more of a collector function while a parallel turnpike and parkways carry through traffic.

The E Branch suffers from considerable headway variation (bunching), causing overcrowding, long waits, and operation irregularities in the downtown subway where the E Branch merges with the B, C, and D Branches. One objective of this study was to see to what extent giving priority to trains that are behind schedule can reduce headway variation. To account for randomness in the subway portion of the E Branch, the base model generates outbound trains emerging from the portal with a headway coefficient of variation (cv) of 0.1; sensitivity runs used cv values of 0.2 and 0.3. Outbound trains were loaded at a dummy stop before the portal so that outbound loads at the portal were consistent with headways (e.g., long headway trains will be more heavily loaded).

The model changes E Branch operations in two ways to make it a better testbed. First, because the corridor is rather short, the

layover at the Heath Street terminus was eliminated, so that trains arriving outbound at Heath Street departed immediately inbound. This change doubles the length over which the route operates without the benefit of dispatching control, allowing a better test of transit priority's impact on irregularity in operations.

Second, the E Branch is unusual in that its peak direction of travel in the morning is outbound. Therefore, the inbound transit passenger volumes were scaled up until the inbound peak volume was 20% greater than the outbound peak volume (both peak volumes occur at the subway portal), making the peak direction of travel the same for private traffic and transit.

TRAFFIC SIGNAL CONTROL STRATEGY

The key features of the traffic signal control algorithm are presented in a flowchart in Figure 2 and are explained in the following subsections. The traffic signal control algorithms were coded in VAP, the flexible traffic signal control language of VISSIM.

Transit Arrival Prediction

For each intersection approach, transit detectors were located at a travel time of 180 to 270 s (two to three signal cycles) upstream, usually at the exit from a far-side stop or from the intersection immediately after a near-side stop. Between the transit detector and the subject intersection lay one or two signalized intersections and one or four transit stops. Transit travel time from detection to the intersection stopline depends on traffic control at intervening intersections, traffic events such as left-turning cars blocking the tracks (Bostonians will recognize this feature of Boston driving), and dwell times at intervening stops. Dwell times, in turn, depend on passenger

boarding and alighting volumes, which depend on headway (time since the last train) at the subject stop and at previous stops, respectively. This model generates boarding passengers randomly following a Poisson process, while alighting passengers are a given fraction of the volume on board. Unit boarding and alighting times are 2.5 and 1.5 s, respectively. At outbound stops, alighting dominates; inbound, boarding dominates.

For each intersection approach, a formula was estimated from simulated historical data to predict travel time from the designated detector to the stopline as a linear function of headway at the detector. This formulation recognizes that trains with longer headways are likely to have greater boarding and alighting volumes at intervening stops. A typical estimation formula (in seconds), for the inbound approach at Intersection RU, was $\text{travel time} = 94.5 + 0.115 \times \text{headway}$ with $R^2 = 0.84$. An arrival time window for the approach was then established as the predicted time ± 20 s. At Intersection SH, where trains turn and consequently are served by a shorter green phase, the window was ± 10 s.

(Reviewers offered two valuable suggestions that deserve further exploration. One is to use an asymmetric time window, with more time allowed after the expected arrival time. This makes

sense because running time distributions tend to be skewed, and because arriving just before a green window is not as bad as arriving just after a green window. The other suggestion is to use continuous detection instead of point detection. With continuous detection, the expected arrival time of the train can be continually refined and the cycle adjusted accordingly.)

Transit Service Cycle Selection

Next, it is decided in which cycle to try to serve the transit vehicle. In general, two options were considered: compressing phases enough to pull a green phase back until it covers the arrival window, and extending phases so that the green phase one cycle earlier covers the arrival window. The selection logic, presented in Figure 3, starts by projecting the future signal state with "ideal" phase lengths continuing from the most recent phase change. Ideal phase lengths are average phase lengths over the last five signal cycles, updated continuously by the signal control program. The current phase, of course, is constrained to be at least its current length plus clearance time. If in the ideal phase lengths projection the arrival window lies entirely

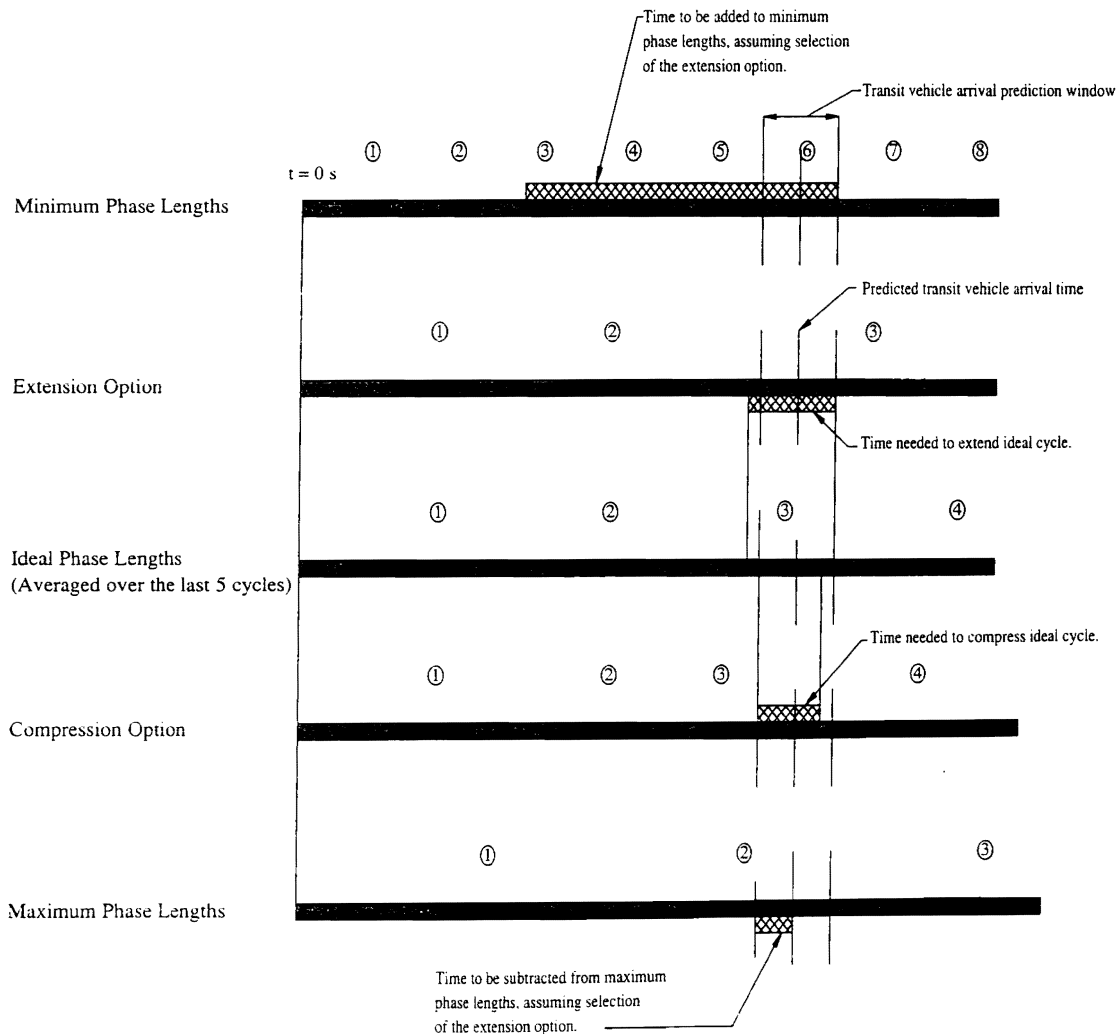


FIGURE 3 Cycle adjustment options.

within a green phase of a particular cycle (e.g., the third cycle), the selection process is over—it is attempted to serve the transit vehicle in that cycle.

Usually, however, ideal phase lengths do not result in the arrival window lying entirely in a green phase. Then the compression and extension options described earlier are created, using the projected ideal phase lengths as a base. The example in Figure 3 indicates a case in which extending phases would allow the train to be served two cycles into the future, while compressing phases would allow the train to be served three cycles into the future.

If both compression and extension options are feasible (respecting minimum and maximum phase length constraints and minimum length of the current phase), cycle selection is based on Δ_{ext} and Δ_{com} , the size of the adjustments needed relative to ideal phase lengths. A penalty function was used that favors extension over compression because of capacity losses that can occur if phases are compressed too much: penalty = Δ_{ext} for extension and $\Delta_{com} + 0.025 \Delta_{com}^3$ for compression. The parameters were selected so that compressing 20 s is penalized twice as much as extending 20 s.

If only one option is feasible, the cycle in that option in which the arrival window lies is selected. If neither compression nor extension can get the entire prediction window to lie in the green, the option with the least expected transit delay is selected, assuming a uniform arrival time distribution across the prediction window. As indicated in Figure 4, if a and b are the amount of the arrival window falling in green and red, respectively, and $c = [(\text{length of red phase}) - b]$ for the extension case only,

$$\text{expected delay} = \frac{b}{b+a} \left(\frac{b}{2} + c \right)$$

Updating Minimum and Maximum Phase Lengths

Once the transit service cycle has been selected, phase lengths and cycle times remain flexible, subject only to minima and maxima that are updated to ensure that it will still be feasible for the arrival window to fall completely in a green phase of the selected service cycle.

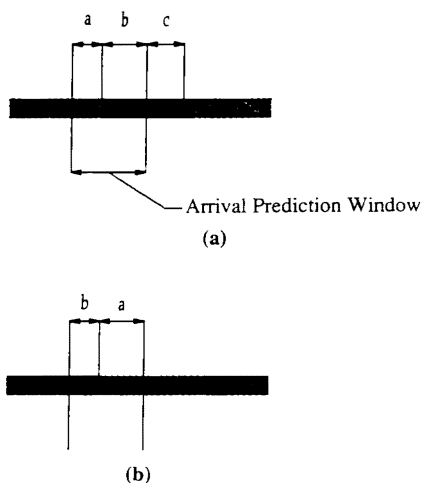


FIGURE 4 Determining expected transit delay: (a) extension option, (b) compression option.

Next, how those phase length limits are updated is described. Referring back to Figure 3, assume that the extension option has been selected, meaning that it has been decided to serve the transit phase in the green period of Cycle 2. First, consider one limiting case, using minimum phase lengths. As indicated in Figure 3, one can readily determine the “additional minimum,” the total time that must be added to minimum phase lengths to make the end of the Cycle 2 green, extended to its maximum length, coincide with the end of the arrival window. Working backward from the phase preceding the selected transit service phase, how much of that additional minimum can be “spent” on each phase, subject to each phase’s maximum length, is calculated. If all the additional minimum can be spent before getting to the current phase, no change is made to the current phase’s minimum time. However, if maximally extending all the intervening phases still leaves some additional minimum to be spent, the minimum length of the current phase is increased by the difference. This process is updated each time the signal phase changes until the green phase of the selected transit service cycle begins. That way, the updated minimum length of whatever phase is current will reflect the actual length of past phases.

An analogous process is used to update maximum phase lengths. This approach preserves maximum flexibility in phase lengths, allowing the signal control program to serve other objectives while still ensuring that the selected green phase will cover the arrival window.

Actuated Control and Coordination

Within the constraints of the updated minimum and maximum phase lengths just described, signal control is free to respond to traffic demands as they are detected. The controller is programmed to respond to three kinds of inputs.

First, a stopline train detector cancels a priority request if the signal is green or yellow. If a train still has not reached the stopline at the end of the arrival window, green is extended 10 s beyond the normal maximum (unless the train’s arrival cancels the request before then). If, in spite of these efforts, a train reaches the stopline on red, it forces intervening phases to their minimum lengths until the train’s phase gets green.

Second, passage and speed detectors located about 60 m upstream of the stopline are used to estimate traffic density with data from the last 3 s (density = flow/speed). When density falls below a minimum, the request from that approach is canceled. A phase will end, subject to minimum phase and clearance lengths, when requests from all its approaches have been canceled. Arrivals during the red period increase the minimum initial green interval by 1 s each up to a maximum of 20 s to protect standing queues that may not extend as far back as the passage detectors.

Third, loose coordination in the inbound direction is provided through peer-to-peer communication, as described by Janos and Furth (5). When a signal’s inbound phase turns green, it transmits the expected arrival time of the platoon to the next inbound intersection. If that intersection is already green, it will hold until the platoon arrives (subject to maximum green). At intersection EV, which has little cross-street traffic, a platoon message from the upstream intersection will force the cross-street green to its minimum length.

Conflicting Transit Requests

If an intersection is timing to meet a train’s priority request when a train from the opposite direction makes a priority request, the

train that is most behind schedule wins and the other's request is canceled.

PRIORITY OPTIONS

Four priority options were tested. One is no priority for transit. In that option, each intersection is fully actuated without a fixed cycle but with peer-to-peer communication providing loose coordination. Because the no-priority option serves as a base case, a fair evaluation of the priority strategy demands a well-formulated no-priority option. The no-priority option performed very well, with average total delay per vehicle only 16.5 s at the average intersection, substantially better than performance of the (primarily fixed-time) control program actually in place. The second option, labeled adaptive, is the traffic control strategy proposed, with every train requesting priority. In the third option, "conditional," only trains that are behind schedule request priority. Conditional or schedule-based priority can be an effective tool for operational control, while limiting the impact on other traffic (1, 6). In the fourth option, labeled "preemption," priority is given to all trains based on stopline detection only. In no case was a train ahead of schedule intentionally held.

Earlier research has emphasized the need for a fine-tuned schedule to apply conditional or schedule-based priority (1). By trial and error and making many simulation runs, a schedule was developed so that, when schedule-based priority was applied, trains would be behind schedule roughly 50% of the time.

EVALUATION AND ANALYSIS

Results are the average of three runs, each simulating morning peak-period conditions for 2.5 h following a 0.5-h warm-up period.

Delay to Transit and Traffic

The overall impact of the four priority strategies can be assessed by examining delay relative to unimpeded operation at each intersection to transit vehicles, to other main-street traffic, and to cross-street traffic. As indicated in Figure 5, adaptive priority reduces transit delay per intersection from 21 to 10 s on average. At one intersection, average transit delay is reduced by 23 s. Under conditional priority, only 43% of the trains approaching intersections were late and thus requested priority; the delay reduction is correspondingly smaller. Preemption did not reduce delay nearly as much as adaptive priority, indicating the benefit of advanced detection.

While priority has a dramatic effect on transit delay, its effect on other traffic is almost negligible. Average delay per vehicle per intersection for main-street and cross-street traffic varies by less than 3 s among the priority options. Pedestrian crossing delay is primarily a function of average cycle length. The priority strategies changed average cycle length by less than 6 s at any intersection, showing no substantial effect on pedestrians, whose needs were intentionally protected by long clearance times and by maximum phase lengths that prevented long cycles. These results indicate the power of a good transit priority algorithm—significant improvement for transit with little impact on other road users.

Transit Travel Time

Transit travel impacts, including both mean and variation in travel time, are presented in Figure 6. Mean travel time impacts parallel the delay impacts, with adaptive priority providing a mean reduction of about 1.5 min per direction. Of perhaps greater importance is the difference in 95th percentile travel time (estimated as mean + 1.64 standard deviations), because transit schedules usually are written to provide enough running time plus layover to account for the 95th percentile running time, and because transit users often believe they must budget enough time for their trip to account for a 95th percentile worst-case performance. Inbound, the improvement compared with no priority is dramatic, with a reduction of 4.4 min for adaptive priority and about 3 min for the other two priority strategies. The differences outbound are also substantial, with absolute priority indicating a reduction of almost 3 min. These reductions are great enough to have potentially significant effects on operating cost and ridership.

Headway Regularity and Crowding Impact

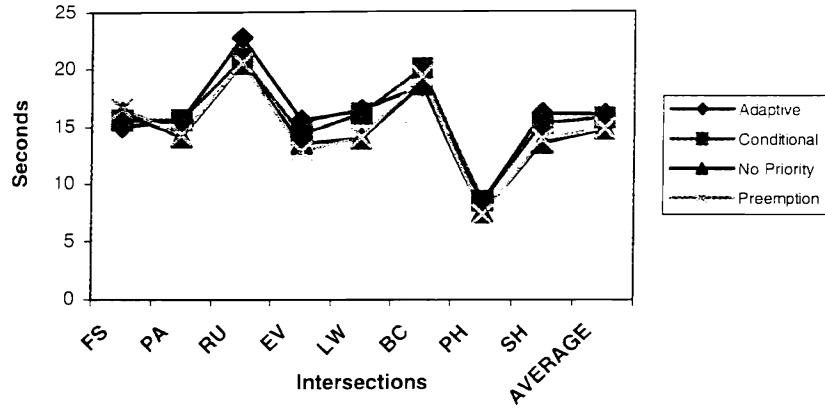
The impact of priority on headway regularity and crowding is presented in Figures 7 and 8. First, the effect of priority on the cv of headway is presented in Figure 7, which indicates how the headway cv changes from when they first emerge from the portal until they return to the portal, using initial cv values of 0.2 and 0.3. Without priority, the classic tendency to bunch is clear from the rise in cv—for example, from 0.3 when emerging from the portal to 0.84 when returning. With conditional priority, the tendency to bunch is still present but is remarkably damped. In this setting, the modest run-time savings, which priority can "push ahead" a train that is late—about 12 s per intersection—is not enough to overcome the self-reinforcing tendency to bunch with high-capacity transit vehicles that begin with significant variations in headway and load.

The primary negative effect of headway irregularity is extreme crowding that can occur on late vehicles. This impact is indicated in Figure 8, where the 95th percentile load at the inbound peak point under the four different priority options is presented. These tests use an initial headway cv of 0.1, and, of course, all cases have the same average load. The impact of priority is substantial, with 95th percentile load at the peak point falling from 173 with no priority to 150 with priority.

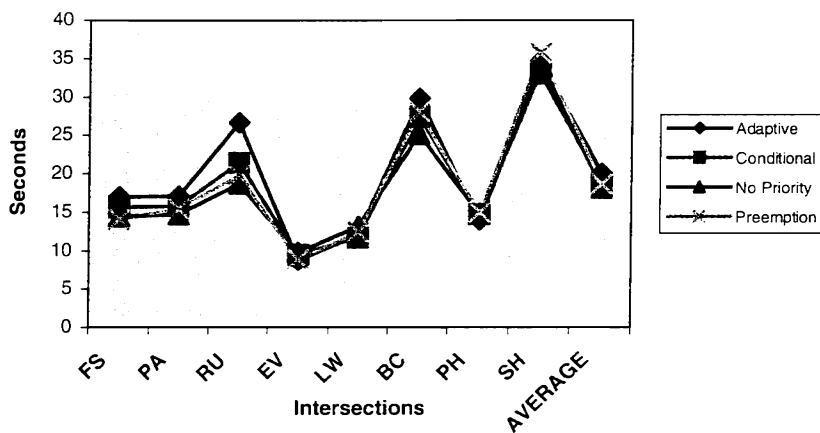
Prediction Success

The success of the control algorithm in predicting train arrival time and in getting trains to arrive on green is indicated in Table 1, which presents the results from the conditional priority scenario. Of the roughly 200 requests for priority, the train arrived at the stopline during the prediction window only about two-thirds of the time. When the train arrived during the prediction window, it always arrived on green. When it did not, it was still possible to have it arrive on green about half the time. Overall, then, the train arrives on green about five-sixths of the time.

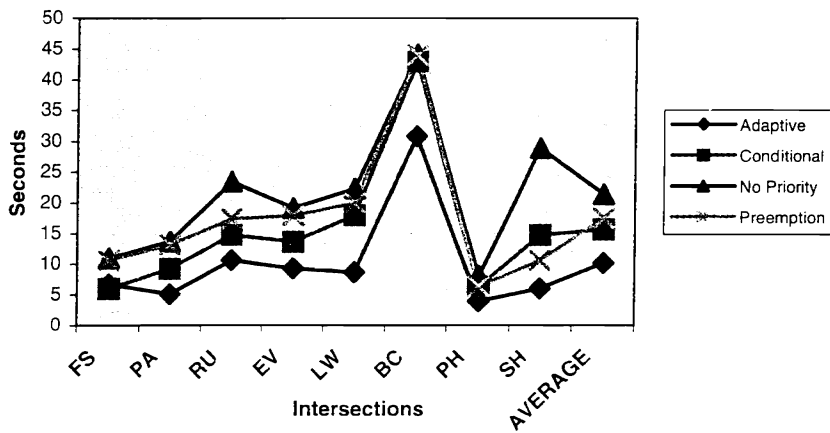
This result indicates both success and potential for improvement with better prediction methods. Another way to improve the strategy is to use interim detectors to update the arrival time prediction and signal cycle adjustment process.



(a)



(b)



(c)

FIGURE 5 Delay impacts: (a) main-street delay; (b) cross-street delay; (c) transit delay.

Results with Greater Traffic and Transit Volumes

Some have said that priority cannot be applied on heavily saturated roads because priority is too disruptive. To show how nondisruptive the adaptive strategy is, corridor operations were modeled with all private traffic volumes (main street and cross

street) doubled. The delay impacts are presented in Figure 9. While doubling traffic volumes naturally increases traffic delay, the effect on private traffic of applying priority is negligible, while transit delay is reduced by 13 s per intersection. When considering the interaction of traffic volume with transit delay, bear in mind that, of the eight signalized intersections, trains operate

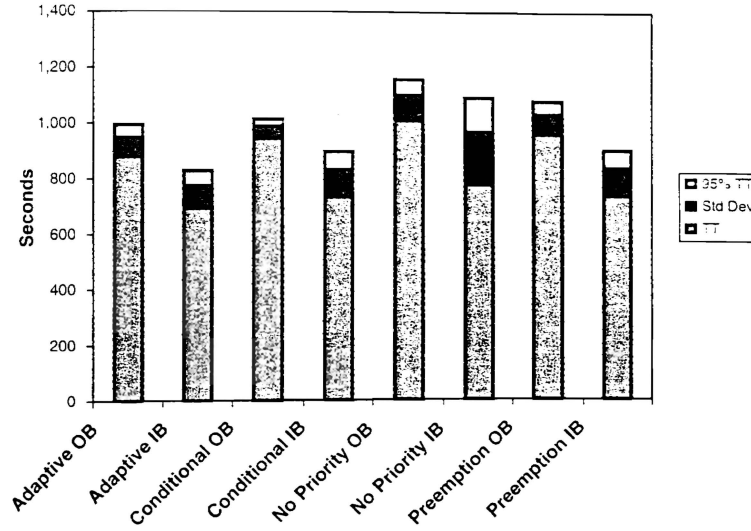


FIGURE 6 Transit travel time impact (TT = travel time, OB = outbound, IB = inbound).

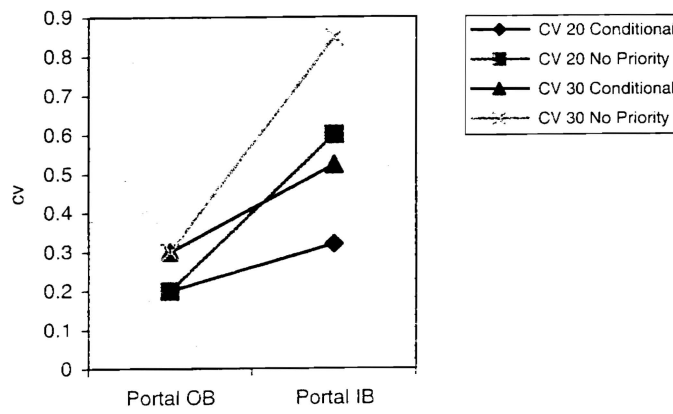


FIGURE 7 Impact on headway cv (OB = outbound, IB = inbound).

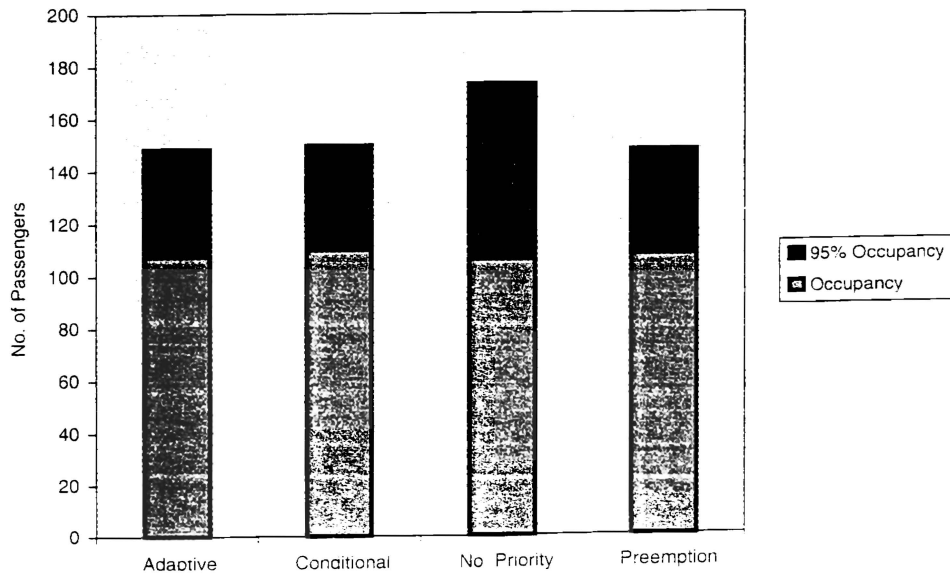


FIGURE 8 Crowding impact.

TABLE 1 Success in Predicting Arrival Time and Adapting Signal Cycle

	Train Arrivals		
	In Green	In Red	Total
In Window	126	0	126
Outside Window	40	37	77
Total	166	37	203

in mixed traffic at intersections SH and PH and at one approach to intersection BC.

Finally, because the impacts of any priority strategy depend partly on how frequently priority requests are made, corridor operations were modeled with transit passenger volumes doubled and train frequency doubled to one train every 4.5 min in each direction. Under this scenario, each intersection sees a train arrival every 2.25 min on average, so that conflicting requests will be frequent, making travel time prediction more difficult, and causing more interruptions to private traffic. The delay impacts are presented in Figure 10. This test underestimates the transit benefit, because it was conducted with arrival time prediction models that had been generated with the original demand levels.

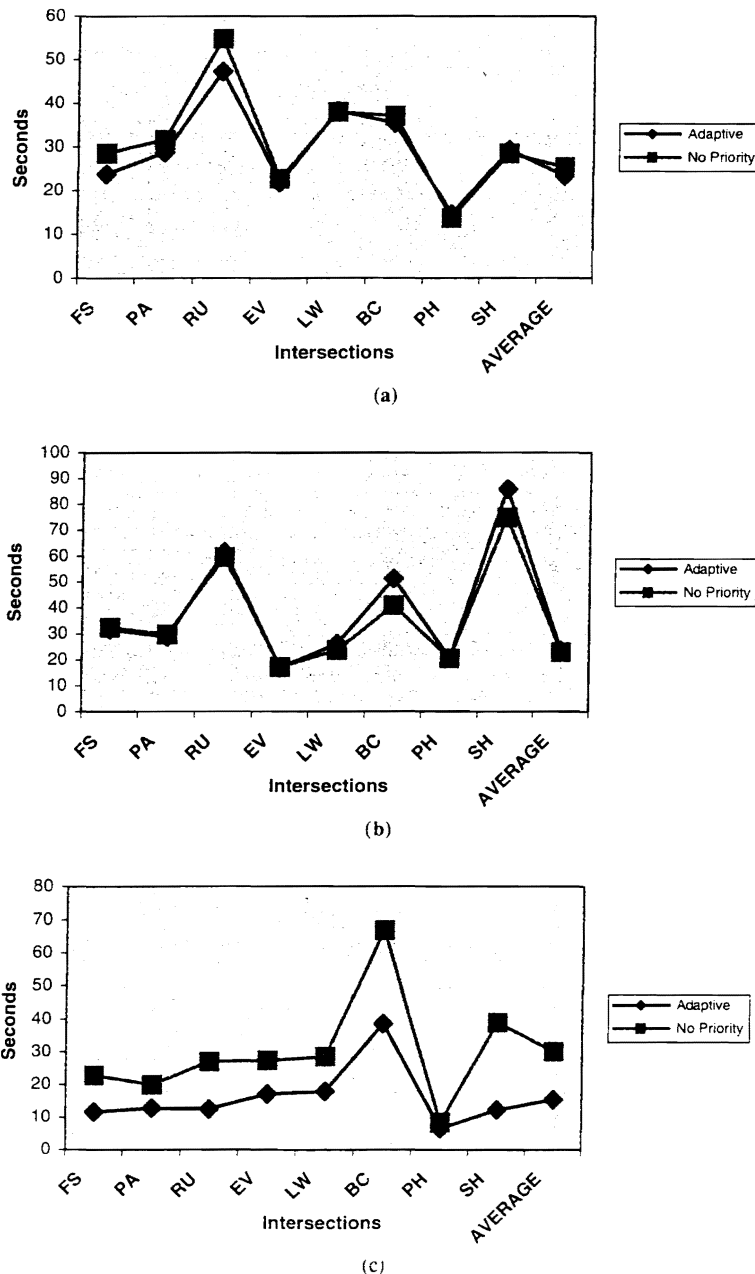
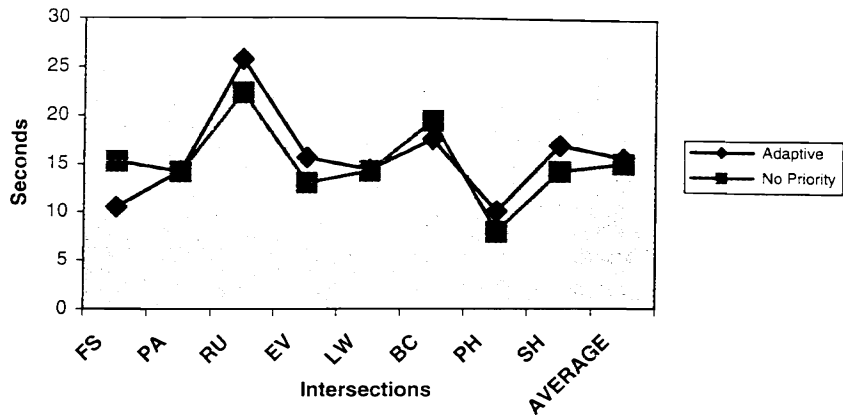
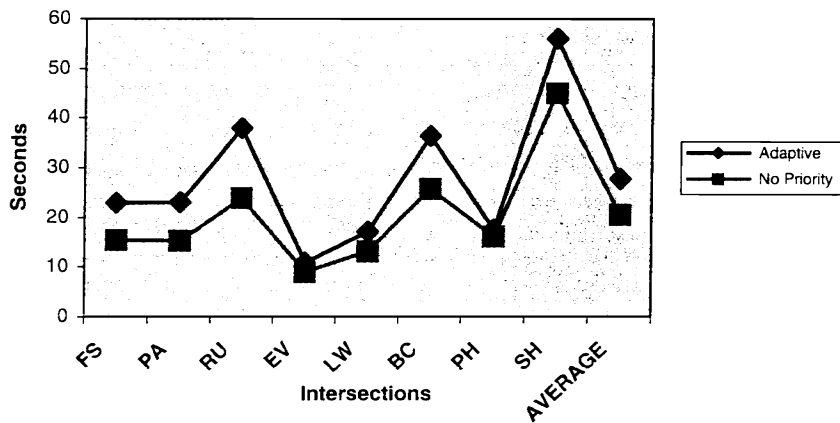


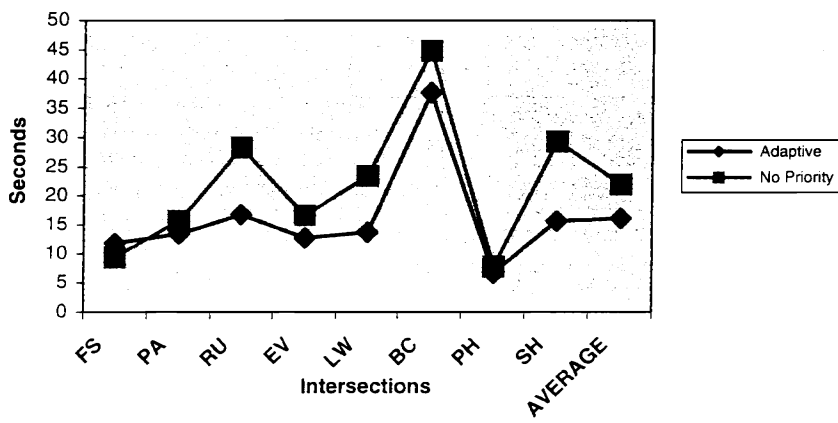
FIGURE 9 Delay impact with doubled traffic volumes: (a) main-street delay; (b) cross-street delay; (c) transit delay.



(a)



(b)



(c)

FIGURE 10 Delay impact with doubled transit volumes: (a) main-street delay; (b) cross-street delay; (c) transit delay.

CONCLUSION

The basic question the authors set out to investigate is whether, in spite of randomness in transit running times, a control strategy based on advanced prediction could provide passage with near-zero delay for transit with little impact on other traffic. While the prediction and control algorithms leave room for improvement in future generations, the results indicate strong support for the adaptive, advanced prediction approach. Transit benefits in terms of reductions in delay, 95th percentile running time, headway irregularity, and crowding are substantial and exceed benefits achieved with near-term prediction only. Effects on other motor traffic are mostly below the threshold of practical (and political) significance. These results support the idea that this approach deserves attention, both for practical implementation and for research to develop improved prediction and adaptation algorithms.

ACKNOWLEDGMENTS

Thanks to Duncan Allen of the Parsons Corporation for graciously providing a working model of Huntington Avenue in VISSIM, to W.K. Law of National Taiwan University for valuable supervision in the early stages of this research while on sabbatical at Northeastern

University, and to Larry Head of Gardner Systems for suggesting the prediction approach that was used.

REFERENCES

1. Furth, P. G., and T. H. J. Muller. Conditional Bus Priority at Signalized Intersections: Better Service Quality with Less Traffic Disruption. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1731*, TRB, National Research Council, Washington, D.C., 2000, pp. 23–30.
2. Head, L. *Improved Traffic Signal Priority for Transit*. TCRP Project A-16 Interim Report, TRB, National Research Council, Washington, D.C., 1998.
3. Langdon, S. M. Simulation of Houston Light Rail Transit Predictive Priority Operation. *ITE Journal*, Vol. 72, No. 11, 2002, pp. 28–32.
4. *VISSIM User Manual, Version 3.60*. PTV Vision, Karlsruhe, Germany, 2000.
5. Janos, M., and P. G. Furth. Bus Priority with Highly Interruptible Traffic Signal Control: Simulation of San Juan's Avenida Ponce de Leon. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1811*, TRB, National Research Council, Washington, D.C., 2002, pp. 157–165.
6. Muller, Th. H. J., and P. G. Furth. Integrating Bus Service Planning with Analysis, Operational Control, and Performance Monitoring. *Proc., Intelligent Transportation Society of America Annual Meeting*, Boston, May 2000. Intelligent Transportation Society of America, Alexandria, Va., 2000.

Publication of this paper sponsored by Committee on Traffic Signal Systems.