Networkwide Approach to Optimal Signal Timing for Integrated Transit Vehicle and Traffic Operations

Matthew I. Koch, Daniel C. Chin, and Richard H. Smith,
Johns Hopkins University

An intermodal networkwide strategy is presented for the determination of optimal traffic signal timings in locations such as reserved transit malls in central business districts where light-rail transit (LRT) vehicles are subject to the same traffic controls as motor vehicles. The determination of optimal signal timings is crucial for public transit since delay at signals contributes significantly to passenger dissatisfaction with the system. Some jurisdictions have little or no coordination of traffic signals with LRT movements, whereas others use some priority systems, often signal preemption, that can seriously disrupt the flow of other traffic at intersections. The strategy considered here is unique because it is both integrated and networkwide, thereby balancing the needs of public transit with those of private vehicles. Traffic and LRT, including all intersections and stations, are treated as one intermodal system for which traffic signal timings are optimized to minimize delay and maximize throughput. The methodology is based on a neural network that determines, in real time, the parameters that control the traffic signal timings. It extends a previously developed methodology, a fundamentally new approach to signal timings for motor vehicular traffic, to the integrated LRT and traffic network. The approach is illustrated by a prototype simulation of part of the Baltimore central business district.

One of the major concerns in the design and operation of light-rail transit (LRT) is the interaction between LRT and other traffic. LRT operations are common on grade level in the central business district (CBD) of many cities, sometimes operating directly on city streets but more frequently using reserved transit malls. Such malls can take several forms. They may consist simply of a portion of a street reserved for LRT with painted lane markings separating the light-rail vehicles (LRVs) from other vehicles, as on Howard Street in Baltimore. Alternatively, such malls may consist of the median strip of a major street (the northern part of First Street in San Jose, for example), or the entire street may serve as a transit mall (the southern part of First Street in San Jose or C Street in San Diego, for example). Although all such malls are successful in keeping LRVs separate from motor vehicle traffic between intersections, the LRVs are subject to the same traffic signal system as other traffic. Although there may be separate phases for LRT at traffic signals, LRVs must often stop for cross traffic. Thus, in addition to the time spent stopping at stations, LRT is subject to significant delays at traffic signals. Because of heavy use of LRT by the general public, such delays significantly reduce person throughput and increase travel time, thereby contributing, in some degree, to public dissatisfaction with public transit. This is a significant problem to address in the LRT planning process.

Approaches to the problem of interaction between LRT and motor vehicles at signalized intersections attempt to give LRT some priority over motor vehicles. There are several methods of giving such priority to LRT, ranging from minor changes in the signal phases to accommodate LRT to more extreme measures such as signal preemption. Although there are many specific implementations of signal preemption, the basic idea is for the LRV approaching a signalized intersection to communicate its proximity to the controller unit at the intersection and then for the controller timer either to extend the green phase to allow the LRV to go through or to force a red phase on cross-street traffic. Either method provides the LRV with a green phase.

There is a good deal of literature on priority systems. For example, some of the considerations involved in the design and planning of priority systems are discussed by Stone and Wild (*1*). One of their main recommendations is that in designing priority systems, particularly signal preemption, total delay should be considered, including both delay for motor vehicle users and delay for LRT passengers. In addition, simulated results showing the effect of preemption on total delay are provided. Radwan and Hwang (*2*) discuss methods of evaluating the preemption system with respect to passenger delay, and some simulation results are given. In addition to the work by Stone and Wild and by Radwan and Hwang, the effects of preemption systems on the throughput and delay of motor traffic are discussed by Gibson et al. (*3*), Celniker and Terry (*4*), and Yagar and Han (*5*). As these authors show, the effect of preemption on the integrated system of traffic and LRT is mixed. Clearly, preemption can speed the passage of LRT, but it can also have unexpected and sometimes adverse effects on cross traffic. In fact, in San Diego [see discussion by Gibson et al. (*3*)], studies conducted soon after the opening of the LRT system in 1981 indicated that preemption had very little negative effect on other traffic, whereas later studies there [see discussion by Celniker and Terry (*4*)] did show a negative impact. These results prompted a change from a preemption system to a more passive system of priority for LRVs.

In this paper a significantly different approach to the interaction of LRT and traffic at signalized intersections along a transit mall is presented. The approach produces, in real time, signal timings that are tuned to approach optimality relative to a measure of effectiveness (MOE) that accounts for both LRT and motor vehicle traffic. The MOE reflects total traffic and transit passenger delay in the system and is calculated with real-time sensor data from both the LRT and traffic components of the system. This approach is very different from an approach based purely on preemption because, from the outset, the signal operation here will be responsive in real time to the needs of the total intermodal traffic and transit system rather than to the needs of transit alone. Although this approach is perhaps most useful on transit malls where preemption is a less popular alternative, it can still be used in conjunction with preemption. For example, at times of heavy traffic, when preemption might cause large queues to build up on cross streets, preemption could be suspended and this approach used. Again it is to be emphasized that this approach is designed to mitigate delay in the combined transit and traffic network in an integrated way, emphasizing the tradeoffs between LRT and motor vehicle traffic.

In this approach, the networkwide traffic control methodology of Spall and Chin (*6*) is combined with a simple but realistic model for LRT movements on reserved malls to obtain a strategy for signal timing control in the integrated system. In the methodology of Spall and Chin, traffic sensor data are used to obtain networkwide signal timings without having to use a detailed flow model for the traffic network. Attempts to obtain reliable models for real-time traffic control on a networkwide basis have been unsuccessful (*6*,p. 1868; *7*,p. 258), largely because of the complexity of vehicular interaction in a large network and the inability to model driver behavior. Thus, an approach not dependent on large-scale models is a desirable feature. The traffic control strategy, as discussed by Spall and Chin, applies to motor vehicle traffic. In order to extend its applicability to the intermodal traffic and LRT system, a model for LRT movements is developed in this paper. In contrast to the difficulty of modeling traffic on a networkwide basis, modeling of LRT, at least on reserved malls as considered in this paper, is simpler. On the reserved malls being considered, interactions with traffic take place only at intersections, and therefore LRT moves with predictable regularity. (In this paper, interaction between LRT and pedestrians is ignored.) Therefore, in contrast to modeling general traffic flows, modeling such LRT movements is a reasonable task.

This paper is organized as follows: The integrated signal control strategy is described, showing how the results of Spall and Chin (*6*) are extended to include LRT. The LRT model, which is critical in such an extension, is discussed in more detail next. Prototype simulation results, showing how the approach could be used on a portion of the Baltimore CBD, and a summary are provided.

Signal Timing Control Strategy

In presenting the signal timing strategy, first a discussion of the strategy as it pertains to motor vehicle traffic alone is given, as it was originally developed by Spall and Chin (*6*). Then its extension to the joint LRT and traffic network is given.

As discussed earlier, the signal timing strategy uses real-time traffic flow data to produce signal timings that are optimized relative to a predetermined MOE. The MOE used here is person delay. Results of simulation studies using this strategy and MOE are provided by Chin and Smith (*8*). However, the approach has the additional advantage that it is readily adaptable to other MOEs. The mathematical techniques are based on use of a neural network (NN) as an approximation to the true, but unknown, mathematical function that controls the signal timings. Performance of the NN depends upon accurate estimation of a large number of parameters called *weights,* which, together with current traffic flow data, determine the signal timings. Accurate estimation of the weights is accomplished by optimization of the MOE using the method of simultaneous perturbation stochastic approximation (SPSA), a general optimization tool well suited for multivariable problems. Optimization of parameters is performed by efficiently extracting information from repeated applications of small, simultaneous perturbations to all the estimated parameters over a period of several days. The timings and associated weights are estimated to apply to a large range of per-cycle traffic fluctuations that vary from light to congested conditions and from smooth to surge traffic behavior. The weights are updated from day to day by SPSA in a gradually adaptive process that proceeds to optimize the signal timings relative to the prescribed MOE calculated from traffic sensor measurements. The real-time traffic flow data are used in two ways. The first is to update the NN weights from one day to the next. The second is to calculate the most appropriate signal timings for the immediate signal cycle using the most recent set of NN weights.

In extending this approach to cover the intermodal system of LRT and traffic, the two main issues are determining what real-time data are to be used and determining what the MOE is to be. Whereas a wealth of real-time traffic data is available from sensors, the same may not be true for LRT. However, it is reasonable to expect that some measurements are available of the times at which the LRV passes a few known points along the transit line. Therefore all that is required are these measurements.

The transit model described in the next section then "fills in" estimates for intermediate times along the route. There are several means of obtaining such measurements. One such way, which is becoming popular in the transit industry, is the Global Positioning System (GPS). Other vehicle location systems, such as scanners on utility poles, would also be adequate for the measurements that are needed in this process. [Several such systems are discussed in Vuchic's text (9, p. 288).] Even with modern automatic vehicle location systems such as GPS, it is still necessary to have a model for LRT for those times when data from the vehicle location system are not available. (This may occur when the GPS signal is blocked by tall buildings, for example.) In addition to time measurements, detailed measurements of passenger loads are useful. In their absence, however, rough, average values could be used as an alternative.

Analogous to traffic flow data, the information provided by the transit model is used in two ways. First, this information is used to update the NN weights from one day to the next and then to actually determine the signal timings at the current time. The latter is a particularly important use for the model because the timings change depending on the presence of an LRV, and it is the model that provides the information about the location of the LRVs.

In order to accommodate LRT as well as traffic, the MOE used in the traffic control strategy is augmented with terms reflecting delay in transit. Delay is with respect to a target schedule, namely, a schedule that could be reached with minimal delay time at traffic signals. Thus, a target schedule is somewhat more optimistic than the usual public transit schedule. To compute the MOE, the procedure is first to determine, at each intersection, whether an LRV will arrive there according to the model. If so, a delay term for the LRV is computed as

$$\text{Delay}_{\text{LRV}}(t) = n(t)[S(t) - M(t)] \tag{1}$$

where

$S(t)$ = time from the point of most recent measurement according to the target schedule,
$M(t)$ = corresponding time as computed by the model, and
$n(t)$ = number of passengers (or average number of passengers if the exact figure is not available).

The delay term in Equation 1 is squared and summed over all LRVs in the network and added to the delay terms for traffic queues. The combined delay expression is then used in the mathematical optimization routine. Similar to the traffic count data used as input by the NN control process as described by Spall and Chin (6), the combined traffic and LRT control process uses these data plus LRT real-time location data to determine signal timings for the next traffic signal cycle.

Time and Location Model for LRT Movements on Reserved Malls

In this section a more detailed description of the model for LRT movements is provided. As stated earlier, this model is a history of locations and corresponding times that the LRV passes each location. The locations are known, the arrival times at one or more previous locations are known exactly, and the model estimates the time component of LRV arrival at each new location.

The time component of the transit model is obtained by estimating the time that the LRV spends at stops, including both traffic signals and stations, and the time it takes for the LRV to travel from one stop to the next. Because only movements on a reserved transit mall (rather than movements on a city street) are considered, it is reasonable to suppose that traffic signal and station stops are the only stops the LRV makes. (The same would be true on a private right-of-way away from the CBD, with even fewer traffic signals.) The procedure is repetitive, starting at a point with an available accurate measurement and estimating along the way until reaching the next point with a new accurate measurement, and then beginning the process again. This leads to a time history at all traffic signal and station stops. The model assumes constant start-up acceleration, constant speed during cruising, and constant braking. Therefore, it is a straightforward matter to compute times at locations intermediate to the stops. The approach is stochastic in nature, and therefore random variability in these physical parameters can be incorporated into the model.

To estimate the time spent in motion from one stop to the next, the regimes-of-motion method discussed by Vuchic (9, pp. 159–174) is followed. When the distance between stops is long enough for the LRV to reach maximum speed, the stop-to-stop time is

$$t_{\text{stop-to-stop}} = \left[\frac{D}{V} + \frac{V}{2}\left(\frac{1}{\bar{a}} + \frac{1}{\bar{b}}\right)\right] \quad (2)$$

where

D = distance (m) between stops,
V = maximum speed (m/sec),
$\bar{a}$ = acceleration rate (m/sec²), and
$\bar{b}$ = braking rate (m/sec²).

When the distance between stops is shorter and the LRV does not reach maximum speed, the stop-to-stop time is

$$t_{\text{stop-to-stop}} = \sqrt{\frac{2(\bar{a} + \bar{b})D}{\bar{a}\bar{b}}} \quad (3)$$

where D is again the distance between stops.

The time t_{stopped} (sec) that the LRV spends at stops, both signals and station stops, as indicated in Figure 1, is now estimated. If the LRV arrives during the red phase with r sec remaining until the next green phase begins, then $t_{\text{stopped}} = r + t_s$, where t_s (sec) is the start-up (or reaction) time, the time it takes the LRV to start once given the green indication. When the LRV stops at a station, the time it spends stopped is the sum of the lag time t_L (sec) from when the vehicle stops to when the doors open; the dwell time t_D (sec), that is, the passenger ser-

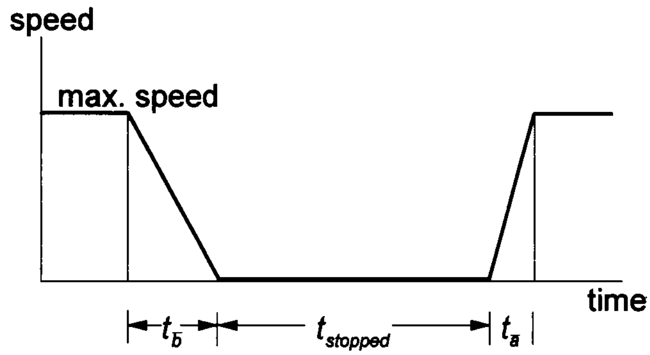

FIGURE 1 Model time profile for LRT stop cycle (does not include start-up time).

vice time; and start-up time t_s, described above. Thus, $t_{\text{stopped}} = t_L + t_D + t_s$. The time spent accelerating $t_{\bar{a}}$ (sec) and the time spent braking $t_{\bar{b}}$ (sec) in Figure 1 are included in the stop-to-stop time discussed earlier. The *Highway Capacity Manual* (10, Chapter 12) provides nominal values for these times, although in practice they can be determined in field tests.

The model for LRT movements is now complete. Starting with a point where an accurate measurement of time is available, the time to all stops and potential stops (i.e., green traffic signals) can be calculated. Equations 2 and 3 are used to estimate the time that the LRV is moving, and the discussion in the previous paragraph is applied to estimate the time the LRV is stationary. Mathematically this can be expressed as follows. If the stops (stations and traffic signals) are denoted as stop (0), stop (1), stop (2), and so on, the time that the LRV arrives at stop (k) is

$$T_{\text{stop}(k)} = \sum_{j=0}^{k-1} \left\{ t_{\text{stop}(j)\text{–to–stop}(j+1)} + t_{\text{stopped[stop }(j)]} \right\}$$

To fill in at other locations between stops, Equations 2 and 3 can also be used. It is important to emphasize that knowing the exact location at all points is not critical because the stochastic (SPSA) nature of the approach allows the control algorithms to accommodate random variations in all parameters in the model. For clarity, terms representing random variations have been omitted from the discussion.

Baltimore Prototype Simulation

The integrated transit and traffic control strategy is now illustrated with a prototype simulation study of a portion of the Baltimore CBD. The configuration of the simulation area is shown in Figure 2. There are 17 signalized intersections with 38 queues, whose timings are controlled in this simulation. Although there are other

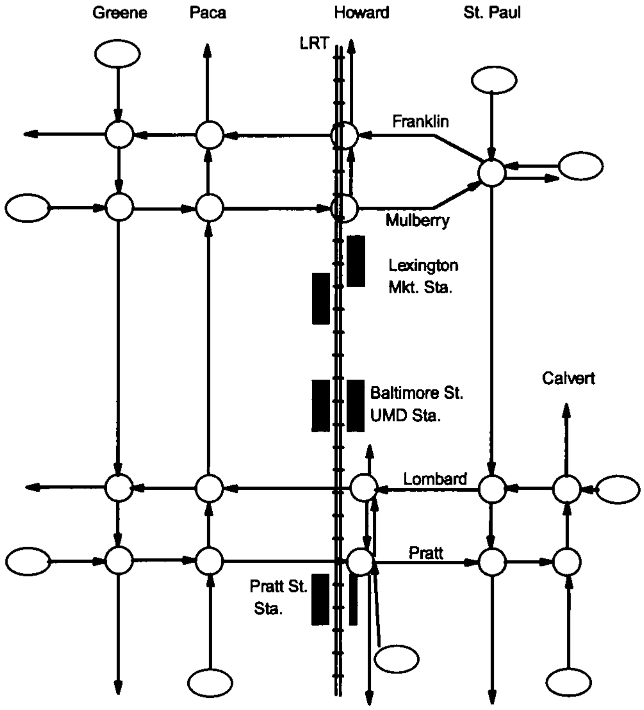

FIGURE 2 Baltimore simulation configuration. Circles denote signalized intersections and ovals denote input nodes.

streets with signalized intersections within the grid, the chosen streets are the major thoroughfares within the Baltimore CBD. Four of these intersections include LRT, which operates double-tracked on Howard Street. Motor vehicle traffic is heavily restricted on Howard Street, and therefore the only traffic flows considered there are those shown in Figure 2. Three LRT stations are shown in Figure 2; actually only the Baltimore Street and Lexington Market stations are within the simulation grid, and the Pratt Street station is on the border of the grid.

The simulation period covers the evening peak period from 4:00 to 6:00 p. m. Traffic flows and initial signal timing information used in the simulation were derived from data supplied by the Baltimore Department of Transportation. Saturation conditions are present on all portions of the east-west streets. Peak-period cycles are 110 sec, and the splits generally favor east-west traffic (*11,* p. 5). In particular, splits on intersections on Howard Street, north of Lombard Street, provide the green signal to east-west traffic about 70 percent of the time. In the simulation (and in practice), LRT trains with three cars operate every 15 min at a maximum speed of about 32 km/hr (20 mph).

The integrated LRT and traffic control strategy was used to determine signal timing splits for the 17 intersections in the simulation network throughout the simulation period. As mentioned earlier, the timing splits are determined by NN weights; for this scenario 1,033 of these weights, with two hidden layers, were estimated. (The greater versatility afforded by two hidden layers as opposed to one is required because the transit and traffic system is not linear.) Although the estimated timings change continuously (cycle-to-cycle) depending on traffic flow and LRV position and load, the underlying NN weights are changed in an optimal adaptive manner over a longer-term basis (days and weeks). In this scenario, this adaptive process lasted about 3 months.

Table 1 shows the average person delay over the 2-hr simulation period at the end of the adaptive process for both the traffic and transit components of the network. Results are shown for three cases: (*a*) fixed signal timings, similar to the current system (baseline); (*b*) timing splits determined by the integrated transit and traffic strategy discussed in this paper, and (*c*) timing splits determined by a simulation of preemption without the integrated strategy. For traffic, person delay is simply the expected time spent waiting at red signals for all motor vehicles in the network times 1.4, which was used as an average value of persons per vehicle. For LRT, delay is as given by Equation 1 relative to a target schedule and reflects the passenger load, consistent with the MOE as discussed earlier. The target schedule is computed using the time and location model and is the fastest possible schedule. (Preemption provides a 10-sec slower schedule since it is assumed that the operator will have to slow down to ensure having the green indication. Therefore, the delay with preemption is small, but not zero.) Passenger load was derived from data supplied by the Mass Transit Administration, Maryland Department of Transportation.

As shown in Table 1, the SPSA-based integrated transit and traffic control strategy is very effective in reducing person delay in the network both for traffic and for LRT.

TABLE 1 Person Delay for Integrated Transit and Traffic Control Strategy Compared with Other Approaches

	Person Delay (hours)		
Method	LRT	Traffic	Total
Fixed Signal Timing (Current Baseline)	42	288	330
LRT Preemption Only	4	318	322
Integrated Transit and Traffic Strategy	33	272	305

When compared with signal preemption, the integrated strategy (and the baseline) is, as expected, less effective than preemption for the LRT portion alone. However, as shown in Table 1, preemption can significantly increase delay to motor vehicle traffic. This increase is 17 percent above the integrated approach and 10 percent above the baseline. When applied to the total network, therefore, preemption shows a 6 percent increase in delay above the integrated approach and slight (2 percent) decrease in delay compared to the baseline. Compared with the current baseline, the integrated approach reduces LRT delay by 21 percent and traffic delay by 6 percent. For the total network, this translates into an 8 percent reduction in delay with the integrated approach.

Summary

A networkwide strategy for the determination of optimal traffic signal timings that balances the needs of LRT and motor vehicle traffic is provided. It operates in real time, using information about the LRV's position and traffic flow data. Through prototype simulations, it showed the capability to significantly decrease delay in the total (LRT plus traffic) network when compared with either the fixed-interval-type controller (currently used) or the preemption method that is popular with several LRT systems.

ACKNOWLEDGMENT

This work was supported by the Independent Research and Development Program, Applied Physics Laboratory, Johns Hopkins University. The authors would like to thank Frank Murphy and Tom Lipko of the Baltimore Department of Transportation and Larry Dougherty of the Maryland Mass Transit Administration for providing data used in the simulation study.

REFERENCES

1. Stone, T. J., and W. A. Wild. Design Consideration for LRT in Existing Medians: Developing Warrants for Priority Treatments. In *Special Report 195,* TRB, National Research Council, Washington, D.C., 1982, pp. 170–175.
2. Radwan, A. E., and K.-P. Hwang. Preferential Control Warrants of Light Rail Transit Movements. In *Transportation Research Record 1010,* TRB, National Research Council, Washington, D.C., 1985, pp. 69–75.
3. Gibson, P. A., B.B. Lin, and R. Robenhymer. Traffic Impacts of Light Rail Transit. In *Special Report 195,* TRB, National Research Council, Washington, D.C., 1982, pp. 161–169.
4. Celniker, S., and E. W. Terry. Trolley Priority on Signalized Arterials in Downtown San Diego. In *Transportation Research Record 1361,* TRB, National Research Council, Washington, D.C., 1992, pp. 184–187.
5. Yagar, S., and B. Han. A Procedure for Real-Time Signal Control That Considers Transit Interference and Priority. *Transportation Research B,* Vol. 28B, No. 4, 1994, pp. 315–331.
6. Spall, J. C., and D.C. Chin. A Model-Free Approach to Optimal Signal Light Timing for System-Wide Traffic Control. *Proc., Conference on Decision and Control,* Institute of Electrical and Electronic Engineers, New York City, 1994, pp. 1868–1875.
7. Newell, G. F. *Theory of Highway Traffic Signals.* Institute of Transportation Studies, University of California, Berkeley, 1989.
8. Chin, D.C., and R. H. Smith. A Traffic Simulation for Mid-Manhattan with Model-Free Adaptive Signal Control. *Proc., Summer Computer Simulation Conference,* Institute of Electrical and Electronic Engineers, New York City, 1994, pp. 296–301.
9. Vuchic, V. R. *Urban Public Transportation: Systems and Technology.* Prentice-Hall, Englewood Cliffs, N.J., 1981.
10. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
11. Sabra, Z. A., and A. Halkias. TRAF-NETSIM: A Practical Tool for Traffic Preemption and Roundabout Intersection Control Modeling. *McTrans,* Vol. 9, No. 2, 1994, pp. 5–6.