Efficient Transit Priority at Intersections

SAM YAGAR

On most transit routes, private vehicles and public transit share a common right-of-way. However, their respective operations are very different from one another, causing an adverse interaction, especially when transit vehicles stop to load and unload passengers on-line at signalized intersections. The severity of traffic delays that are caused when transit operations are ignored by traffic signal control models is illustrated. The impacts on traffic flow caused by transit vehicles stopped to load passengers on-line are illustrated in terms of a typical arrival profile at an intersection, including both cars and a streetcar. It is seen that the streetcar loading operation can significantly reduce capacity and cause delay to both transit and private vehicles, especially when the signal optimization does not take this phenomenon into account. It is shown that, by considering transit loading effects when designing signal timings, delays to both transit and private vehicles can be reduced. Fixed- and real-time methods for providing appropriate transit priority to reduce travel times for transit passengers, and sometimes also to private vehicles, are discussed.

Public transit is used in large metropolitan areas to move large numbers of people to and from the city center without severely affecting the limited urban road capacity. On the one hand, the relatively large buses and streetcars help to achieve this, whereas on the other hand, the nonhomogeneity of operation that they introduce into the traffic operation can disturb the traffic flow.

On most transit routes, private vehicles and public transit share a common right-of-way. However, their respective operations are very different from one another, causing an adverse interaction, especially when transit vehicles stop to load and unload passengers on-line at signalized intersections. Although this interaction is not easily modeled, it cannot be ignored when modeling the operation for traffic signal optimization (*1*). The effects of the traffic impedance caused by the on-line transit loading are bad enough when the loading process is accounted for by the models but can be much worse if the models fail to recognize the loading process, because they then optimize for a pseudooperation, and the signal timings can be meaningless. Therefore, the University of Waterloo is developing models to capture the effects of this transit loading procedure for both fixed- and real-time signal operation. For fixed-time control, some transit modeling enhancements are being made to upgrade the TRANSYT-7F model so that it will represent transit loading phenomena appropriately and thus produce reasonably efficient signal timings (*2*). Also, in the interim, whereas existing operational models cannot properly optimize mixed public and private operation, the Metropolitan Toronto Transportation Department (Metro) and the Toronto Transit Commission (TTC) have been cooperating (*3*) in an attempt to improve the overall people-moving capability of the road system in terms of capacity and delay by use of real-time signal preemption on Queen Street. Metro and TTC are implementing a system of real-time transit priority for streetcars on the Queen Street corridor, and the University of Waterloo is also developing a real-time traffic signal optimization model that is sensitive to transit effects and can give priority to transit when this is desired and appropriate (*4*).

The following sections illustrate the effects of on-line transit loading and briefly describe fixed- and real-time approaches that are being suggested for developing efficient signal timings with due consideration for transit in terms of (*a*) modeling the traffic operations and (*b*) giving appropriate weights to the transit vehicles so that the operator can consciously attempt to minimize either vehicle delay or person delay.

THE QUEEN STREET EXAMPLE

Attempts to apply the bus provisions in the state-of-the-art TRANSYT-7F model to optimize the fixed-time signal operation in Toronto proved unsuccessful. Whereas TRANSYT-7F claims to consider transit effects, it cannot represent the traffic blockage caused when vehicles load on the traveled way. The effects of this will be shown in the section on fixed-time procedures. While stopped, buses cause varying amounts of delay, and streetcars can virtually close an approach if they load from the sidewalk (as is the case at most signalized intersections on Queen Street). Whereas the immediate effects are felt by private vehicles, the capacity reduction is usually also felt by transit vehicles and passengers. The wasted capacity can cause queue buildups, which affect all subsequent vehicles arriving on the shared approaches. These adverse impacts are exacerbated when the optimization models used to select signal timings are unaware of the blockages caused by transit.

The intersection of Queen and Bathurst streets is shown in Figure 1. This intersection serves a demand of 90 streetcars per hour during the peak period, split almost equally among the four approaches. The streetcar and private traffic volumes are shown in Figure 1. With the exception of the northbound approach on Bathurst, there is no refuge from traffic for passengers to access the streetcar. They must walk between the sidewalk and the streetcar, which loads in the median lane. While passengers are getting on and off the streetcar, all traffic in the approach must wait, causing full blockage of the approach.

MODIFICATIONS TO FIXED-TIME PROCEDURES

The TRANSYT-7F model claims to be able to represent the mixed operation of transit vehicles in the traffic stream using

---

Department of Civil Engineering, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.

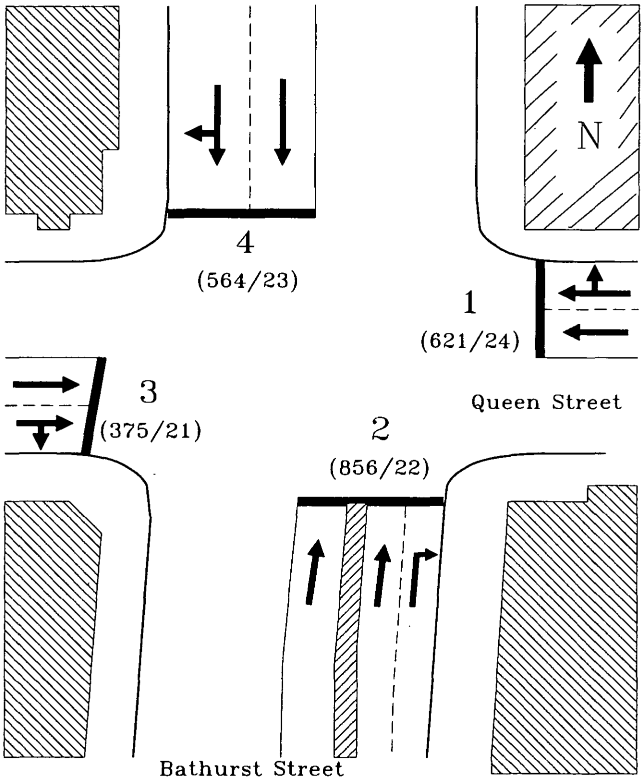

FIGURE 1 Movements and traffic volumes (cars/streetcars) on the four approaches.

a technique introduced in Britain into the TRANSYT/5 version (5). However, this transit provision is not appropriate to normal North American operating conditions. As discussed earlier, when the transit vehicle loads in the traveled right-of-way, it blocks some or all of the road. This is especially critical when a streetcar in the median lane loads passengers from the sidewalk at a signalized intersection, as is the case in the Queen Street corridor in Toronto.

Transit Representation by Current TRANSYT-7F Model

The top portions of Figures 2, 3, and 4 show a typical arrival flow profile for cars in one direction on one approach to a hypothetical intersection if the intersections are closely spaced so that there is no platoon dispersion. In Figures 2, 3, and 4 the normal flow profiles above the time axis represent individual cars *a*, *b*, . . ., *u* in order of arrival, with either 0, 1, or 2 cars moving in each 2-sec time slice. The double X (one above the other) represents the arrival of a streetcar. This arrival profile is assumed to repeat every cycle, in this case every 60 sec.

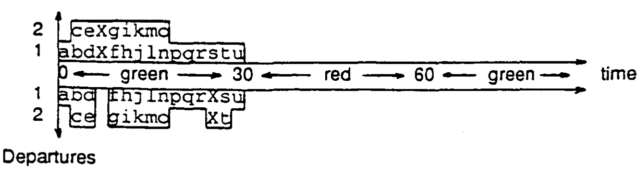

FIGURE 2 Vehicle arrivals and departures for off-line loading of streetcar.

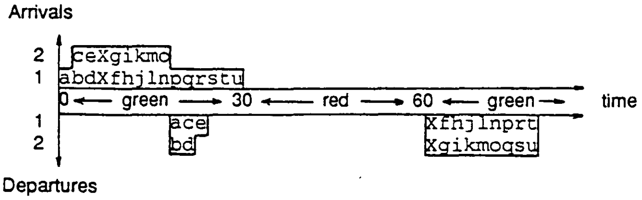

FIGURE 3 Vehicle arrivals and departures when signal settings do not consider streetcar loading effects.

The flow profiles below the time axis represent departures, whose maximum rate of two vehicles per 2-sec time slice represents a saturation flow of 3,600 vehicles per hour of green. For purposes of comparison, Figures 2, 3, and 4 all show the departure profiles for the same vehicles that arrived between $t = 0$ and $t = 30$. Also, Table 1 gives the cumulative departures from $t = 0$ beginning with the same vehicle *a*, which arrived between $t = 0$ and $t = 2$ sec.

For simplicity we suppose that saturation flow equals 3,600 vehicles per hour of green, so that each 2-sec period serves up to two vehicles, and that start-up loss is 1 sec. The typical TRANSYT-type arrival pattern of Figure 2 has 1, 2, 2, 0, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, and 1 cars arriving in the successive 2-sec periods, for a total of 21 per cycle. In Figures 2, 3, and 4 and Table 1 the streetcars arrive at 8 sec, 68 sec, 128 sec, and so forth. A streetcar displaces about two cars in terms of the saturation flow of the intersection approach.

The simplest way to illustrate and discuss the effects of transit interference is to assume that a streetcar arrives in every cycle at the same relative position within the cycle, as in the top portions of Figures 2, 3, and 4, and that the time taken to load is always the same. Figures 2, 3, and 4 assume a 60-sec cycle (30 green and 30 amber + red) and use the same 18-sec effective loading time.

The departure flow profiles shown on the lower portions of Figures 2, 3, and 4 represent the following cases:

- Figure 2 shows the case where streetcars load off-line and do not hold up private vehicles (cars).
- Figure 3 shows the equilibrium flow profiles that would result when a streetcar loads on-line in each cycle, but TRANSYT-7F sets the traffic signals as if they loaded off-line, as in Figure 2.

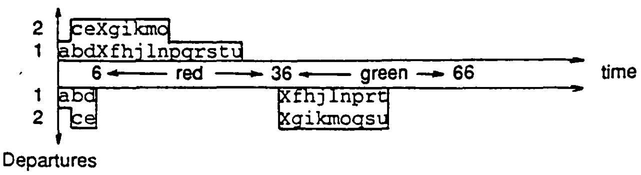

FIGURE 4 Vehicle arrivals and departures when signals are set in response to streetcar loading effects.

TABLE 1 Numbers of Vehicle Arrivals and Departures in Successive 2-sec Intervals

Time (secs)	No. of Arrivals		Cumulative Car Departures for (Fig. 2)	Cumulative Car Departures for (Fig. 3)	Cumulative Car Departures for (Fig. 4)
	str	car			
0		0	0	0	0
2		1	1	0	1
4		3	3	0	3
6		5	5	0	5
8	1	5	5	0	5
10		7	7	0	5
12		9	9	0	5
14		11	11	0	5
16		13	13	0	5
18		15	15	0	5
20		16	16	2	5
22		17	17	4	5
24		18	18	5	5
26		19	18	5	5
28		20	20	5	5
30		21	21	5	5
32		21	21	5	5
34		21	21	5	5
36		21	21	5	5
38		21	21	5	5
40		21	21	5	7
42		21	21	5	9
44		21	21	5	11
46		21	21	5	13
48		21	21	5	15
50		21	21	5	17
52		21	21	5	19
54		21	21	5	21
56		21	21	5	21
58		21	21	5	21
60		21	21	5	21
62		22	22	5	22
64		24	24	7	24
66		26	26	9	26
68	2	26	26	11	26
70		28	28	13	26
72		30	30	15	26
74		32	32	17	26
76		34	34	19	26
78		36	36	21	26
80		37	37	21	26

- Figure 4 shows the equilibrium flow profiles when streetcars load on-line, but this is recognized and taken into account when setting the signal timings.

The flow profiles of Figures 2, 3, and 4 are discussed below.

Streetcar Loads Off-Line (Figure 2)

The signal offset is set to accommodate cars *a*, *b*, . . ., *u* with perfect progression, as TRANSYT-7F would strive to do. TRANSYT-7F would turn the signal green at times 0, 60, 120, and so forth, if other network conditions did not mitigate against this. TRANSYT-7F's timings would not be affected significantly by a loading transit vehicle, since TRANSYT-7F assumes that the transit vehicle travels and loads on a parallel link, entering the shared right-of-way only to preserve its relative position in queue for the signal. Since we have perfect coordination, the vehicles merely pass through the intersection except for the streetcar, which loads off-line from $t = 8$ sec to $t = 24$ sec, at which time it leaves, as shown by the double X. The streetcar takes up both units of capacity and causes car *s* to be delayed and leave with car *t*.

Effects of Transit Loading on the Right-of-Way (Figure 3)

The lower portion of Figure 3 shows what happens to the departure pattern when the signal timings are developed under the incorrect assumption that the transit loading operation does not affect the flow profile. This is explained as follows:

1. The streetcar arrives at 8 sec (upper diagram) but finds a queue in front of it.
2. As will be confirmed by the calculations in Steps 4 and 5 following, the queue in front of the streetcar is not served

until $t = 24$ (after car e has left). The streetcar begins to load at this time and finishes at $t = 42$ (i.e., during the red phase).

3. It then waits for the next green and leaves at $t = 60$ followed by the vehicles that arrived behind it ($f, g, \ldots, u$, with u leaving at $t = 76$ as shown on the bottom portion of Figure 3). Therefore, arrivals $A, B, \ldots, E$ of the next cycle, which arrive between $t = 60$ and $t = 66$ joining the queue behind vehicle u, leave between $t = 78$ and $t = 84$. (The vehicles that arrive in the next cycle, 60 sec after $a, b, \ldots, e$, are labeled using capital letters $A, B, \ldots, E$, respectively.)
4. Since the process is cyclical, vehicles $a, b, \ldots, e$, which arrived one cycle (60 sec) earlier (between $t = 0$ and $t = 6$) would also depart 60 sec earlier than vehicles $A, B, \ldots, E$ of the next cycle (between $t = 18$ and $t = 24$) as shown on the bottom portion of Figure 3.
5. The streetcar that arrived at $t = 8$ would reach the front of the queue after vehicle e departs. This confirms that the streetcar can indeed start loading at $t = 24$.

The equilibrium pattern has each streetcar arriving 8 sec into the green, waiting in queue for 16 sec, then loading into the red phase and leaving at the beginning of the next green.

Recognition and Accommodation of Transit Vehicles (Figure 4)

In Figure 3 we can see that each streetcar queues for 16 sec and then loads for a further 18 sec, thus holding up other traffic for the 18-sec period. If TRANSYT-7F could see this, and if other networkwide factors did not dictate otherwise, TRANSYT-7F would want to turn the signal red at times 8, 68, 128, and so forth so that the streetcars could arrive on a green phase and begin to load immediately but not hold up any other traffic while loading. This would result in the output patterns shown in Figure 4. It would save the streetcar 16 sec of queuing time and reduce the delays to private traffic caused by the loading. Table 1 gives the cumulative vehicle arrivals and the calculated cumulative car departures corresponding to arrivals after $t = 0$. The values given in Table 1 are as follows:

Column	*Value*
1	Time
2	Cumulative streetcar arrivals
3	Cumulative car arrivals
4	Cumulative car departures if the streetcar loads off-line
5	Cumulative car departures if the streetcar loads on-line
6	Cumulative car departures if the traffic signal is adjusted to accommodate the on-line transit loading

The greater the number of vehicles departing at a given time, the better the given tabulated system works. When the cumulative number of vehicles leaving equals 21, all of the vehicles that arrived in the first cycle (between $t = 0$ and $t = 30$) have been served. For the purposes of this illustration, Table 1 was tabulated up to 80 sec, which is enough to show when cumulative departures have reached at least 21 for all cases. Columns 4 and 6 show that some of the second cycle arrivals (after $t = 60$) have already departed by $t = 80$, as indicated in the operations of Figures 2 and 4, respectively.

Discussion of Signal Plans

Off-line loading gives the minimum delay, as would be expected. The earliest departure profiles are in Figure 2, and in Table 1 the highest numbers of departed vehicles at any given time are in Column 4. The same timings would give much poorer performance if the streetcar loads on-line, as is seen in Figure 3 and Column 5 of Table 1. However, if we know the effects on traffic caused by the streetcar loading on-line, we can take this into consideration in setting the signals. The result would resemble Figure 4 and Column 6 of Table 1.

The delays shown in Figure 3 and attributed to TRANSYT-7F are probably realistic, or at least unbiased, representations of how this intersection would perform in a network whose signals were timed using TRANSYT-7F, in view of the fact that TRANSYT-7F does not fully represent the interaction between transit and private vehicles. However, we confess that within a network context we could not likely optimize a given intersection to perform as in Figure 4. Therefore the above estimated saving is really an upper bound. We need a computer model such as TRANSYT-7F to facilitate the analysis of large networks, but the model that is used must recognize the important interactions between transit and private vehicles and be able to represent these effects in its optimization routines. We were able to represent the effects of a streetcar loading in the shared right-of-way with the use of dummy preemptive signals and parallel subnetworks (*6*).

We are currently addressing this model development issue with emphasis on the key problem of transit arrivals in some cycles and not others and the corresponding effects on the periodicity that the TRANSYT model assumes. This involves an adaptation of the TRANSYT-7F model to the more difficult situation of nonperiodic arrivals of streetcars in some cycles and not others (i.e., treating a nonstationary problem with a basically stationary model) (*2*).

REAL-TIME MODELING

In theory, real-time models can accommodate the noncyclical effects of transit vehicles loading on-line more readily than fixed-time models. However, real-time models are relatively new, and have seen very few applications in North America to date. Some of these are described briefly below.

Current Applications in North America

Real-time control is being tested in Canada by applying the SCOOT (*7*) model in Red Deer and Toronto. However, the transit-related problems of TRANSYT-7F also affect SCOOT: it does not recognize the loading effects of transit, and it is basically an evolutionary model of TRANSYT-7F plans that does not respond quickly enough to treat the effects of loading transit vehicles, which occur in some cycles and not others.

OPAC (*8*), a responsive model that uses dynamic programming, is being developed in the United States. However, it is difficult to model transit loading in this type of optimization due to the much larger set of suboptimal states that would have to be considered and stored at each optimization stage.

SCAT (*9*), a real-time model developed in Australia, is being used in a 28-signal network in Oakland County, Michigan. There are plans to add about 80 more signals to the SCAT system. SCAT does not consider mixed transit/traffic operation.

Representing Transit Effects and Providing for Transit Priority

As we said before, Metro and TTC are testing real-time priority on the Queen Street corridor. At the same time the Signal Priority Procedure for Optimization in Real-Time (SPPORT) real-time model (*4*) is being developed to provide real-time signal control under such conditions. It incorporates traffic-responsive signal control methods and takes into account the effects of transit vehicles on traffic flow. Included are facilities to simulate and evaluate its own operation.

Whereas the current version of SPPORT examines only individual intersections, future versions are planned to establish integrated systems of isolated intersections sharing advance information for coordinated real-time network control. For now, it is considered a reasonable approximation to treat intersections with large uncoordinated traffic volumes on competing approaches as isolated intersections.

Development of Signal Timings

SPPORT requires one or more lists of important events or activities, ordered by priority, to which it responds in allocating green time. The higher on the list, the more likely an activity/event is to receive a green phase when requested by the occurrence of that type of event. If there is only one absolute prioritized list of activities/events, SPPORT merely generates the timing sequences rigidly according to detected activities/events, as a preprogrammed traffic cop might do.

However, SPPORT can use the high-speed capability of a computer to generate alternative signal timings and provide respective local optimum solutions for consideration by its own simulation and optimization routines. It generates timing sequences and preevaluates the corresponding traffic operations according to each of any number of alternative priority lists. Each list can be considered as representing the relative priorities accorded by a different traffic expert or traffic cop. Each list has a different order for the events, reflecting its own unique set of relative priorities.

These distinct lists are used to generate alternative traffic signal timing sequences for a time horizon equal to that for which there is advance information on traffic demands. SPPORT preevaluates each of the timing sequences generated from the respective priority lists and dynamically selects the most promising timing plan on-line for immediate short-term application. It then implements the best plan for a renewal period of typically about 5 sec. Then the whole process rolls over for this typically 5-sec period, renewing itself over and over every 5 sec.

The following is an example list of types of events ordered by priority for the simplified case where there are no buses:

1. A streetcar on the main street–peak direction,
2. Serving a queue on the main street–peak direction,
3. A streetcar on the cross street,
4. Serving a queue on the cross street,
5. A streetcar in the main street–off-peak direction,
6. A queue request from the main street–peak direction,
7. Serving a queue in the main street–off-peak direction, and
8. A queue request from the cross street.

SPPORT is traffic-responsive in that it continually detects and uses traffic information to update the current signal plan. This signal plan update is performed approximately every 5 sec, and the SPPORT system is said to function in real-time because it can perform the update within this 5-sec time frame. This system allows for various levels of transit priority (i.e., transit events can be placed at different levels on the priority lists and transit vehicles can be weighted to reflect their occupancy).

Except by direct user request, SPPORT does not give uncontested priority to transit vehicles (i.e., green extension and red truncation are not used to unconditionally favor transit vehicles at the intersection), because this strategy delays private vehicles and can also delay transit vehicles in the long run. SPPORT's method of comparing various schemes for traffic-responsive signal control allows it to give appropriate priority to transit vehicles without hindering the overall performance at the intersection.

For comparing the timings produced by the respective priority lists to determine the most promising signal plan, SPPORT can use any cost function that is given. These are determined by policy. The following are some possible policies that might be considered: minimum vehicle delay, minimum person delay, and total cost (including person delay and operating costs).

Initial tests using an earlier version of the SPPORT model (*4*) have indicated that real-time traffic-responsive transit-priority traffic signal control could be effective.

Representation and Interpretation of Detector Data

Vehicle detectors allow SPPORT to predict vehicle arrivals at the next detector or at the intersection (Figure 5). SPPORT makes such predictions using the detection time, the estimated speed of the vehicle, the distance between the detector most recently activated and the next detector, and the distance between the detector most recently activated and the intersection. For example, a detector installed 500 m upstream of the intersection can provide between 30 and 50 sec of advanced flow information (at an average traffic speed of about 50 km/hr).

When a detector senses a vehicle, it records two pieces of information: the vehicle type (transit, private, or emergency) and the time at which it detected the vehicle.

Representation of the Traffic Interactions of Transit or Emergency Vehicles

Since transit vehicles hold up other traffic while they load and unload passengers, it is necessary to model their operation in the traffic stream. This is discussed below, and SPPORT's methods for representing these effects are described.

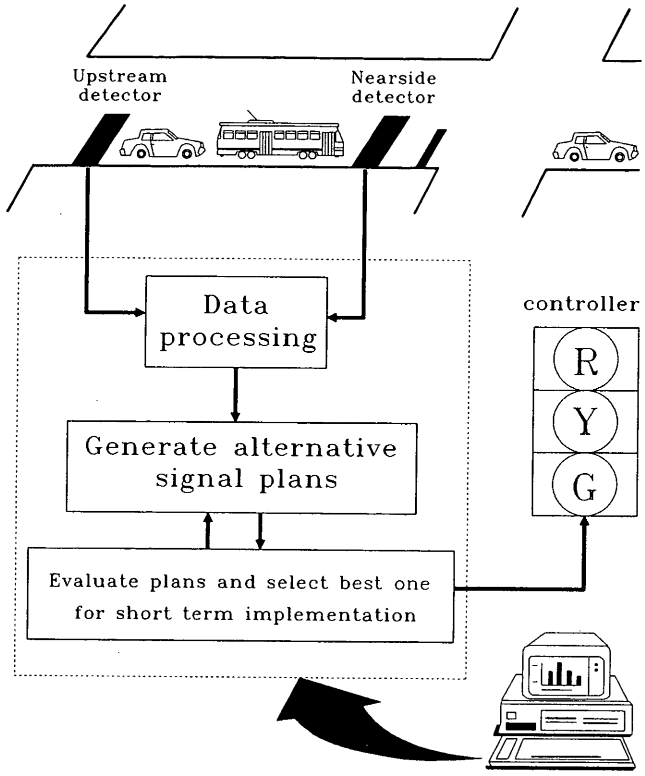

FIGURE 5 Control structure.

Typically, a streetcar holds up traffic on all lanes of an approach while loading and unloading, even when the traffic signal is green. This is modeled by reducing the saturation flow to zero while the streetcar loads. A bus blocks the lane in which it is stopped and may disturb traffic in other lanes, especially when loading near an intersection. These disruptions are modeled in SPPORT by temporarily increasing service headways for approaches on which the transit vehicles are loading.

SPPORT estimates the time of departure from the intersection by using a FIFO (first in, first out) queuing model. Departure time is calculated using the status of the traffic signals and the service headway for the approach on which the vehicle is traveling. Service headways are calculated directly from saturation flows.

The user provides saturation flows for each approach, for each type of vehicle, for each of the following situations that may apply at that approach:

1. No transit vehicles are loading,
2. A bus is loading,
3. A streetcar is loading and blocks only part of the approach,
4. Items 2 and 3 both occur, and
5. A streetcar blocks the whole approach (saturation flow = 0).

The time that it takes for transit vehicles to load and unload passengers at a stop is called the dwell time. The user provides SPPORT with a representative (average or median) dwell time for both streetcars and buses.

The appearance of an emergency vehicle greatly disrupts traffic. It is difficult to predict the resultant flow precisely, because individual responses to the approaching emergency vehicle vary. As an initial investigation, an emergency vehicle has been modeled as a nonstop, very high-priority vehicle. The necessary time taken to clear the queue in front of the emergency vehicle is calculated so that the signal can be turned green at an appropriate time in advance of the arrival of the emergency vehicle at the intersection.

Preliminary tests at the critical intersection of Queen and Bathurst streets in Toronto indicate that SPPORT can reduce delays compared with the current fixed-time control and that transit priority measures can reduce person delay even further (*4*).

CONCLUSIONS

Appropriate consideration and modeling of transit operations as they affect traffic flow is critical to providing efficient signal timings. This is especially critical when transit loads passengers on-line at signalized intersections. The use of appropriate models could improve the productivity of an intersection by increasing its throughput and by decreasing total person hours of delay in traffic, compared with the commonly used fixed-time and real-time control models, both of which fail to represent on-line transit loading.

The emphasis of this paper has been on (*a*) describing the adverse effects on traffic caused by transit vehicles loading on-line and (*b*) outlining methods for efficient management of integrated urban traffic systems with transit vehicles that load on the traveled way at signalized intersections.

ACKNOWLEDGMENT

This work was aided by a grant from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

1. S. Yagar. Accommodating Transit in TRANSYT. In *Transportation Research Record 1181,* TRB, National Research Council, Washington, D.C., 1989, pp. 68–76.
2. M. A. P. Jacques and S. Yagar. Alterations to the TRANSYT-7F Model To Represent Near-Side Transit Stops. Submitted for publication.
3. *Mainline Traffic Signal Priority Phase V—Demonstration Project: Draft Final Report.* Cansult Engineering Ltd., Feb. 1991.
4. S. Yagar, B. Han, and J. Greenough. Real-Time Signal Control for Mixed Traffic and Transit Based on Priority Rules. In *Traffic Management* (Yagar, Courage, and Rowe), Eng. Foundation Press, 1992.
5. J. R. Pierce and K. Wood. *BUS TRANSYT—A User's Guide.* Department of the Environment, TRRL Supplementary Report 266. Crowthorne, England, 1977.
6. P. G. Joyce and S. Yagar. Representing Stochastic Transit Dwell Times in Traffic Signal Optimization. *Transportation Research A,* Vol. 24A, No. 2, 1990, pp. 87–98.
7. P. B. Hunt, D. I. Robertson, R. D. Bretherton, and R. I. Winton. *SCOOT—A Traffic Responsive Method for Coordinating Signals.* LR 1014, TRRL, Crowthorne, England.
8. N. H. Gartner. OPAC: A Demand-Responsive Strategy for Traffic Signal Control. In *Transportation Research Record 906,* TRB, National Research Council, Washington, D.C., 1983, pp. 75–81.
9. J. Y. K. Luk. Two Traffic-Responsive Area Traffic Control Methods: SCAT and SCOOT. In *Transportation Research Record 881,* TRB, National Research Council, Washington, D.C., 1984.

Publication of this paper sponsored by Committee on Bus Transit Systems.