Simultaneous Optimization of Offsets, Splits, and Cycle Time

Nathan H. Gartner, John D. C. Little, and Henry Gabbay, Operations Research Center, Massachusetts Institute of Technology

Setting traffic signals in an urban street network involves the determination of cycle time, splits of green time, and offsets. All existing methods use a sequential procedure for calculating the traffic control variables. A common cycle time is established first, then green splits are calculated for each intersection, and, finally, offsets among the signals are determined. Because these three computation stages are not really independent, the results are often not optimal. This paper describes a new computer method, mixed-integer traffic optimization, designed to optimize simultaneously all the traffic control variables of the network. The method has been programmed in conjunction with the mixed-integer routine of IBM's MPSX optimization system and thus provides a globally optimal procedure. It was applied to several traffic signal networks and is shown to offer certain advantages over existing methods. An example is presented in the paper to illustrate the input requirements, output format, and application of the program.

The traffic control variables in a signalized street network are cycle time, green splits at each intersection, and offsets among the signals. The usual approach for determining these variables consists of three stages.

1. A common cycle time for the network is selected. The signals are then said to be synchronized. This has been shown to improve traffic flow in the network by exploiting the platooned structure of vehicular movement on the signalized links. Because the capacity of a signalized intersection is a function of the cycle length, the common cycle is usually determined by the most heavily loaded intersection, called the critical intersection.
2. Green splits are calculated separately for each signal by apportioning green times for conflicting approaches in proportion to the respective traffic loads.
3. The signals are coordinated by establishing a set of offsets that determine the relative timing among the signals. Only at this stage is a network optimization procedure commonly used.

However, inasmuch as the three stages of signal setting are not independent, this approach cannot guarantee optimality, i.e., the best possible settings for the network. Some of the methods iterate among the different stages to find an improved solution.

Among existing methods, SIGOP (1) scans a predetermined number of cycle times. For each cycle, splits are determined locally at each intersection and offsets are optimized by the OPTIMIZ subroutine by using a random search technique. Network performance is then evaluated in terms of delays and stops, by a coarse simulation of traffic flow on all the links of the network by the VALUAT subroutine. The results obtained from VALUAT dictate the selection of a set of cycle time, splits, and offsets. Three deficiencies of this procedure are apparent. First, the offset optimization procedure determines a local, not necessarily global, optimum; second, the splits are determined independently from the other variables; third, the stochastic nature of traffic flow, which can have a decisive effect on link performance, is not considered. Thus, in recent comparative evaluations of SIGOP and TRANSYT, the lower bound on cycle time was consistently selected as the best value by SIGOP; including stochastic effects would quite possibly have moved the results upward (2, 3, 4).

The combination method and TRANSYT use a critical intersection approach for determining the network cycle time (5). TRANSYT also allows for evaluation of performance at different cycle times, but in view of its rather formidable computing requirements this possibility is seldom used in practice (4). Because capacity of an intersection increases with cycle length, the critical intersection approach is based on analyzing the capacity requirements of the most heavily loaded intersection, i.e., the intersection with the highest sum of representative flow-capacity ratios on all signal phases. The network cycle time is then calculated for this intersection by a method such as Webster's (6) or by specifying the maximal degree of saturation on all approaches to the intersection. This procedure may not be optimal in a network situation because the interaction between the flows and the spatial road network structure of the area is disregarded. Splits and offsets are calculated differently by the two methods. The combination method calculates splits locally at each intersection and uses a

Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.

series-parallel combination procedure, based on link performance functions, for setting offsets. The procedure can be generalized to non-series-parallel networks by means of dynamic programming (7). TRANSYT (8) uses simulation of traffic flow throughout the network to calculate both splits and offsets. Optimization is performed by a hill-climbing method using a one-at-a-time variable search that requires a complete network simulation pass at each step.

Although these systematic methods have been found to achieve significant improvements in traffic performance (9, 10), they cannot guarantee an optimal solution for all the network control variables in a reasonable amount of computer time. Therefore, further improvements in traffic performance can be made by devising a method that simultaneously optimizes all the network variables.

This paper describes a new computer method and program, mixed-integer traffic optimization (MITROP), designed to achieve this goal. The program was developed as part of the urban traffic control research project sponsored by the Federal Highway Administration (11). The approach taken has been to develop a suitable network traffic model and performance measure and then to apply contemporary optimization techniques to it. The principal technique used is mixed-integer linear programming. A number of computer codes are available for this technique; the one used here was the IBM mathematical programming system (12). This paper discusses the main features of the program and presents a detailed example to illustrate the input data requirements and the output format of the program. A more comprehensive description of the program as well as documentation and application to several test networks can be found elsewhere (13).

TRAFFIC SIGNAL NETWORK OPTIMIZATION PROBLEM

The street network consists of nodes and links. The nodes represent the signalized intersections, and the links represent sections of street that carry traffic in one direction between two adjacent intersections. All parameters and variables of the system are defined below, and the traffic signal setting problem is presented as a simultaneous optimization program.

Definitions

Let S_j denote the traffic signal at node j, and let (i, j) denote the link connecting nodes i and j. We define

$r_{ij}(g_{ij})$ = effective red (green) time at S_j facing (i, j);
$\psi_j(AB)$ = intranode offset at S_j, measured as the time from the beginning of green on phase A to the beginning of green on phase B; and
C = cycle time.

The relations between physical and effective signal timings are shown in Figure 1 for a single two-phase signal; they follow the basic model of traffic signal operation used by Webster and others. The signal phases are denoted by single capital letters A, B, and so on, which are replaced by letter pairs, e.g., (i, j), when they are assigned to links. Vehicle platoons released at the start of green at node i travel to node j on link (i, j). Figure 2 shows the fundamental relationship between travel time and internode offset on a link. Let

τ_{ij} = travel time of platoon's head from i to j;
γ_{ij} = arrival time of platoon's head at j, measured relative to start of g_{ij}, so that $-r_{ij} \le \gamma_{ij} \le g_{ij}$; and
ϕ_{ij} = internode offset in extended form.

$$\phi_{ij} = \tau_{ij} - \gamma_{ij} \tag{1}$$

θ_{ij}, the internode offset in reduced form or $[\phi_{ij}]_{\text{mod } C}$ is the time from the start of green at S_i to the start of green at S_j occurring next, so that $0 \le \theta_{ij} < C$. This time is used to calculate the physical settings for the controller.

Objective Function

The objective of the network optimization procedure is to determine signal settings (offsets, splits, and cycle time) that minimize the disutility encountered by vehicles traveling through the signalized intersections. The particular type of disutility can be set by the traffic engineer and may include delays, stops, acceleration noise, or some combination of these measures. The most widely used measures of performance are delays and stops. Because of the inherent fluctuations in the traffic flow process, which induce random variations about the mean in variables such as total flow and temporal distribution of arrivals within a cycle, it is common to separate the total disutility or performance function into two components. The first component is associated with the mean of the traffic flow process. This component is represented in MITROP by the link performance function (LPF). The second component is associated with the random variations about the mean and is represented by the saturation deterrence function (SDF). Our goal is to minimize the total disutility in the network D. Therefore, the objective function is

$$\text{Min } D = \min \sum_{(i,j)} [(LPF)_{ij} + (SDF)_{ij}] \tag{2}$$

LPF and SDF are aggregated over all links (i, j) in the network, including both internal and input links. In general,

$$(LPF)_{ij} = f_{ij} z_{ij}(\phi_{ij}, r_{ij}, C) \tag{3}$$

$$(SDF)_{ij} = Q_{ij}(r_{ij}, C) \tag{4}$$

where,

f_{ij} = average flow on link (i, j) in vehicles/h.
z_{ij} = average loss per vehicle on link (i, j) for traveling through the signal at node j (delay, stops); z_{ij} is a function of all the signal-control variables, i.e., offset, split, and cycle time.
Q_{ij} = the average overflow queue at the signal stop line on link (i, j); that is, the average number of vehicles that because of the random fluctuations are unable to clear the intersection during the cycle in which they arrive. Q_{ij} turns out to be a function of split and cycle time, but not of offset.

The objective function D represents a loss rate in the network such as vehicle-hours per hour, vehicle-seconds per hour, or vehicle stops per hour. A more detailed description, as well as comparison with field data, is given subsequently.

Constraints

The constraints of the optimization program must represent the street network structure as well as all important relationships among the decision variables of the program. For each link in the system we have the following relationship:

$$r_{ij} + g_{ij} = C \quad (5)$$

Equation 5 also implies the relationships between settings on opposing links, involving the lost times for each phase (Figure 1). If the signal is to provide sufficient capacity on each link, then

$$s_{ij} g_{ij} \geq f_{ij} C \quad (6)$$

i.e., the flow arriving during one cycle must not exceed the available capacity during that cycle (s_{ij} denotes the saturation flow rate on the link). To facilitate pedestrian crossing we have a minimum red requirement, $r_{ij} \geq (r_{ij})_{min}$. Because there is little gain in capacity with very long cycle times, we should have an upper limit on the cycle length C_{max}. A maximum cycle length also prevents drivers from becoming impatient or believing that the signals are defective. For safety reasons as well as capacity requirements, it is also desirable to have a lower limit on the cycle time C_{min}.

An important physical contraint in any synchronized

Figure 1. Relation between physical and effective signal timings of basic model for traffic signal operation.

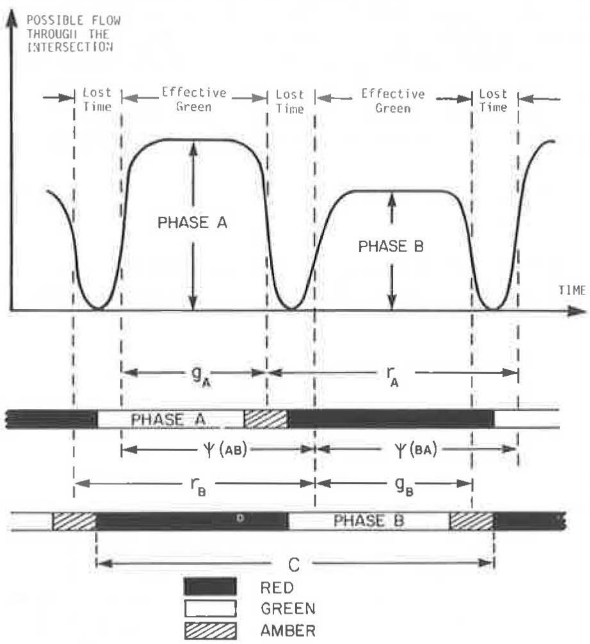


Figure 2. Travel time and offsets on a signal-controlled link.

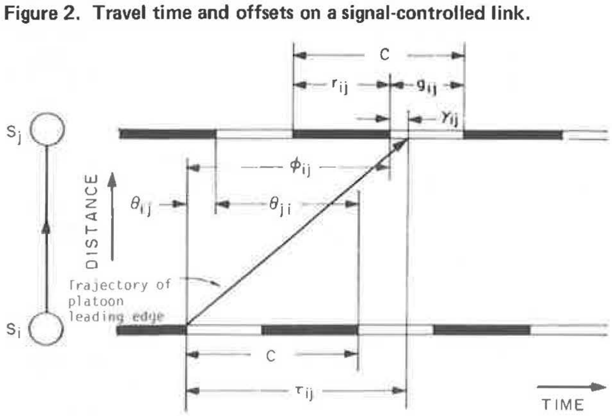


Figure 3. Traffic transition process through a link's exit signal.

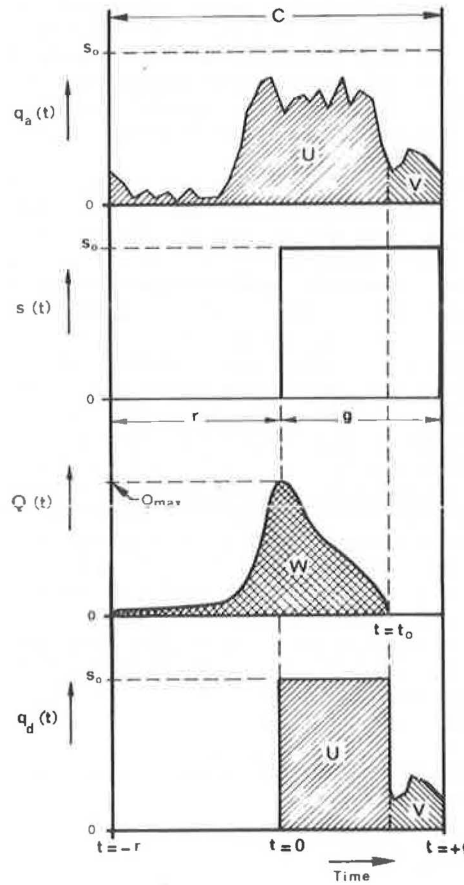
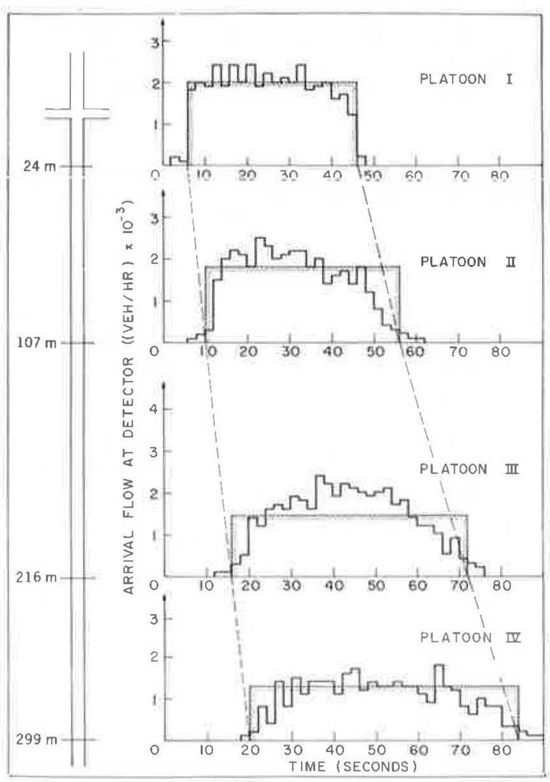


Figure 4. Platoon flow and rectangular approximation at four observation points along a signal-controlled link.



traffic signal network is that the algebraic sum of the offsets around any closed loop of the network equals an integer multiple of the cycle time (14). Both internode and intranode offsets have to be included. In mathematical form, the constraint reads

$$\sum_{(i,j)\epsilon l} \phi_{ij} + \sum_{j\epsilon l} \psi_j(l) = n_l C \quad \text{for each loop } l \tag{7}$$

where n_l = an integer associated with loop l.

Optimization Program

We are now in a position to formulate the traffic signal network optimization problem as the following optimization program: Find values of ϕ_{ij}, r_{ij}, C to

$$\text{Min} \sum_{(i,j)} [f_{ij}z_{ij}(\phi_{ij}, r_{ij}, C) + Q_{ij}(r_{ij}, C)] \tag{8}$$

subject to

$$\sum_{(i,j)\epsilon l} \phi_{ij} + \sum_{j\epsilon l} \psi_j(l) = n_l C \quad \text{for each loop } l \tag{9}$$

$$r_{ij} + g_{ij} = C \quad \text{for each link } (i, j) \tag{10}$$

$$g_{ij}s_{ij} \geq f_{ij}C \quad \text{for each link } (i, j) \tag{11}$$

$$r_{ij} \geq (r_{ij})_{min} \quad \text{for each link } (i, j) \tag{12}$$

$$C_{min} \leq C \leq C_{max} \tag{13}$$

r_{ij}, $g_{ij} \geq 0$; ϕ_{ij}, n_l are unrestricted in sign.

The MITROP processor linearizes in pieces the nonlinear components of the objective function so that the program can be solved by mixed-integer linear programming. The MPSX system then uses branch-and-bound techniques to determine simultaneously the optimal values for the continuous signal-control variables and the associated loop integer values n_l.

LINK PERFORMANCE FUNCTION

This section describes the transition process of traffic through a link and the computational procedure for determining link performance. The beginning of green time at S_j (Figure 2) is established as a reference point. Thus, a cycle period (-r, g) consists of an effective red period (-r, 0) and an effective green period (0, g). We use the following notation:

$q_a(t)$, $q_d(t)$ = arrival, departure rate in vehicles/s,
$A(t)$, $D(t)$ = cumulative number of arrivals, departures at time t during a cycle, and
s = saturation flow rate during the green period in vehicles/s.

From the beginning of any red period at S_j, the following relations exist:

$$A(t) = \int_{-r}^{t} q_a(\tau)d\tau \qquad D(t) = \int_{-r}^{t} q_d(\tau)d\tau \tag{14}$$

If the signal is undersaturated, all vehicles arriving during a cycle in which the red period precedes the green can be accommodated in that cycle. Therefore, all performance calculations for the link can be confined to a single interval (-r, g). The queue length Q(t) is given by the difference between the cumulative number of arrivals and the cumulative number of departures:

$$\begin{aligned} Q(t) &= A(t) - D(t) = A(t) \quad &\text{if } -r < t \leq 0 \\ &= A(t) - ts \quad &\text{if } 0 < t \leq t_0 \\ &= 0 \quad &\text{if } t_0 < t \leq g \end{aligned} \tag{15}$$

t_0 denotes the queue clearance time and is obtained by solving

$$Q(t_0) = A(t_0) - t_0 s = 0 \tag{16}$$

The departure rate is described by

$$\begin{aligned} q_d(t) &= 0 \quad &\text{if } -r < t \leq 0 \\ &= s \quad &\text{if } 0 < t \leq t_0 \\ &= q_a(t) \quad &\text{if } t_0 < t \leq g \end{aligned} \tag{17}$$

The complete transition process is shown in Figure 3. This basic model will now be used to calculate the LPF in terms of the delay encountered by the vehicles. Other performance measures such as number of stops, acceleration noise, or the closely related measure of energy consumption can be similarly calculated. For a general approach to calculating these measures of performance see, for example, Huddart or Chung and Gartner (15, 16).

The delay incurred by Q(t) queuing vehicles during an interval dt is Q(t)dt. Therefore, the total delay time Z incurred by traffic during a full cycle (-r, g) is represented by the area under the queue length curve, i.e.,

$$Z(\gamma, r) = \int_{-r}^{g} Q(t)dt = \int_{-r}^{t_0} Q(t)dt \tag{18}$$

The size of this area depends on the arrival time γ of the platoon of vehicles at the signal and, through equation 1, on the offset ϕ and the split r (and, indirectly, on the cycle time C). The average delay per vehicle z is obtained by dividing by the total number of arrivals during one cycle, A_c, which can be calculated by equation 14.

$$z(\gamma, r) = Z(\gamma, r)/A_C \tag{19}$$

MITROP uses a simple model to calculate the LPF. Traffic flow is represented by a rectangular platoon, which is generally made up of a primary component and a secondary component. Alternative assumptions, such as a tadpoling flow pattern (17), could equally well be made. The platoon must correspond to the average flow, f, on the link. Dispersion effects are taken into account via a platoon dispersion factor, which is a function of the link's length and which may be calibrated for each link. Because this traffic flow model is a simplification of reality, an assessment was made of its quality for calculating delay. Comparisons were made between delay computed from the model and that of actual platoons as reported by Hillier and Rothery (18). In their observations, individual vehicle arrivals were recorded at four locations downstream of a signal-controlled intersection operating on a 90-s cycle. The data were subsequently averaged to give the mean number of arrivals by 2-s intervals at each location. The results are shown in Figure 4. The spreading, or dispersion, that takes place along the approximately 305 m (1000 ft) of roadway downstream of the intersection can readily be seen; e.g., the platoon has a 72-s passage time at the fourth location as compared to the 40-s effective green time at the intersection.

For the comparison, observed platoons were approximated by rectangular platoons containing the same number of vehicles. Dispersion was represented by a piecewise linear function of distance. This yielded the platoons seen superimposed on the observed data of Figure 4. Equations 14 through 19 were then used to calculate delay as a function of arrival time γ for both observed platoons and the rectangular approximations. Various splits were chosen to cover a wide range of possible degrees of saturation of the signals at the different locations. A sample of results of the comparisons is shown in Figure 5. A more extensive evaluation is given elsewhere (13). In most cases we find a close fit between the two functions. The largest deviations occur at extremes of splits or arrival times (and hence offsets), which are expected only rarely in practice. In some cases, the fit, can be improved through better selection of the parameters for the rectangular platoon approximating the actual field data. To be used in MITROP the LPFs are piecewise linearized as shown by the dotted lines in Figure 5. It is noteworthy that in certain regions the linearized approximation is closer to the function derived from the actual platoon than to the function obtained from the rectangular platoon.

SATURATION DETERRENCE FUNCTION

The LPFs are calculated based on the assumption that the traffic flow patterns are identical during each cycle. In practice there are fluctuations about the mean in variables such as total vehicle flow and the temporal distribution of vehicle arrivals within a cycle because of variations in driving speeds, marginal friction, and turns. These random variations can lead to temporary overflow queues that seriously degrade performance. Although this effect is negligible at low degrees of saturation, its predominance at high values has been established in several studies (18, 19, 20). A representation of this effect is needed to prevent green time from approaching its lower bound too closely, thus leading to saturation. If cycle time is also a variable, it is particularly essential that it be determined by the optimization process. This is because there is a fundamental trade-off between capacity loss at short cycles and the inherently large delays of long cycles.

The value of the expected overflow queue, based on the capacity of the signal's approach and the degree of saturation, has been calculated by Wormleighton (21). Following field studies in Toronto, he developed a model describing traffic behavior along a signalized link as a nonhomogeneous Poisson process with a periodic intensity function. According to our notation, the signal's capacity K, in vehicles per cycle, is

$$K = sg \tag{20}$$

and the degree of saturation x is

$$x = fC/sg \tag{21}$$

The results of the computation are given in Table 1. A typical relationship between expected overflow queue and split time in this model is shown in Figure 6. Rather similar deterrence functions have been given by Webster (6) for setting signals at a single intersection and by Robertson (9) for the TRANSYT signal setting model. The rate of delay incurred by the queued vehicles is simply $Q_{ij}(r_{ij}, C)$. This term is called the link's saturation deterrence function and provides the second component of the network objective function given in equation 2. To represent the SDF in a form amenable to mixed-integer linear programming, MITROP makes it piecewise linear, as was done with the LPF. To ensure a minimum level of service, r is restricted so that the degree of saturation stays below 0.95. Thus we obtain an upper limit on the red split, which appears as a vertical constraining line in the (Q, r) diagram. Two additional lines are determined by the two pairs of points (P_1, P_2) and (P_3, P_4), having degrees of saturation $x_1 = 0.95$, $x_2 = 0.90$ and $x_3 = 0.85$, $x_4 = 0.70$ respectively (Figure 6).

SIGNAL NETWORK OPTIMIZATION EXAMPLE

The following data describe a signalized street network for which optimal settings are determined by MITROP by using the IBM 370/165 computer system. Figure 7 shows a sketch of the test network. There are 9 intersections (nodes) and 24 links of which 16 are internal to the network (i.e., interconnect a pair of nodes) and 8 are input links. Table 2 gives the data for the network. Traffic flow on the input links is assumed to be continuous (though random fluctuations are taken into account), which therefore yields a platoon length of one complete cycle time. Table 3 gives the loops of the network; only one independent set of loops has to be considered. Each loop is specified in terms of its links and nodes. For each link (i, j) there is a corresponding internode offset ϕ_{ij} and for each node j there is a corresponding intranode offset ψ_j. These offsets are arranged into constraints according to equation 7, there being a total of eight. In case the orientations of a loop, as given in Table 3, and a link in that loop do not coincide, a negative sign has to be taken for that offset. Table 3 also gives the set of integers, n_l, that has been determined by the optimal solution for each loop l in the network. Table 4 gives the main output data for the network. The optimal value for the objective function includes both deterministic and stochastic delays, according to equation 2.

ANALYSIS OF RESULTS

Experience of researchers and practitioners in urban traffic control has shown that the cycle time may well be the most important of the signal-control variables. This is suggested by both U.K. and U.S. studies (5, 9). The cycle time provides for the necessary capacity to serve the traffic demand at each intersection, and it is the prime determinant of the possible coordination strategy among the signals in the network. Using Webster's notation, we have at each node j in the network the following relationship:

$$\sum_i g_i = C - L_j \tag{22}$$

i.e., the sum of effective green times on all phases i equals the net green time available for crossing the intersection (cycle time minus lost time). Because L_j, the total lost time at node j, is a fixed quantity, the net capacity increases with cycle length. This exposes a trade-off that is amenable to optimization. On the one hand, an increase in cycle time usually increases red times on individual phases and, consequently, the mean waiting time as expressed by LPF. On the other hand, an increase in cycle time also increases capacity and thus reduces stochastic delay due to the randomness in arrivals of vehicles, as expressed by SDF. The interplay between LPF and SDF as a function of cycle time

Figure 5. Comparisons of link performance function for actual and rectangular platoons.

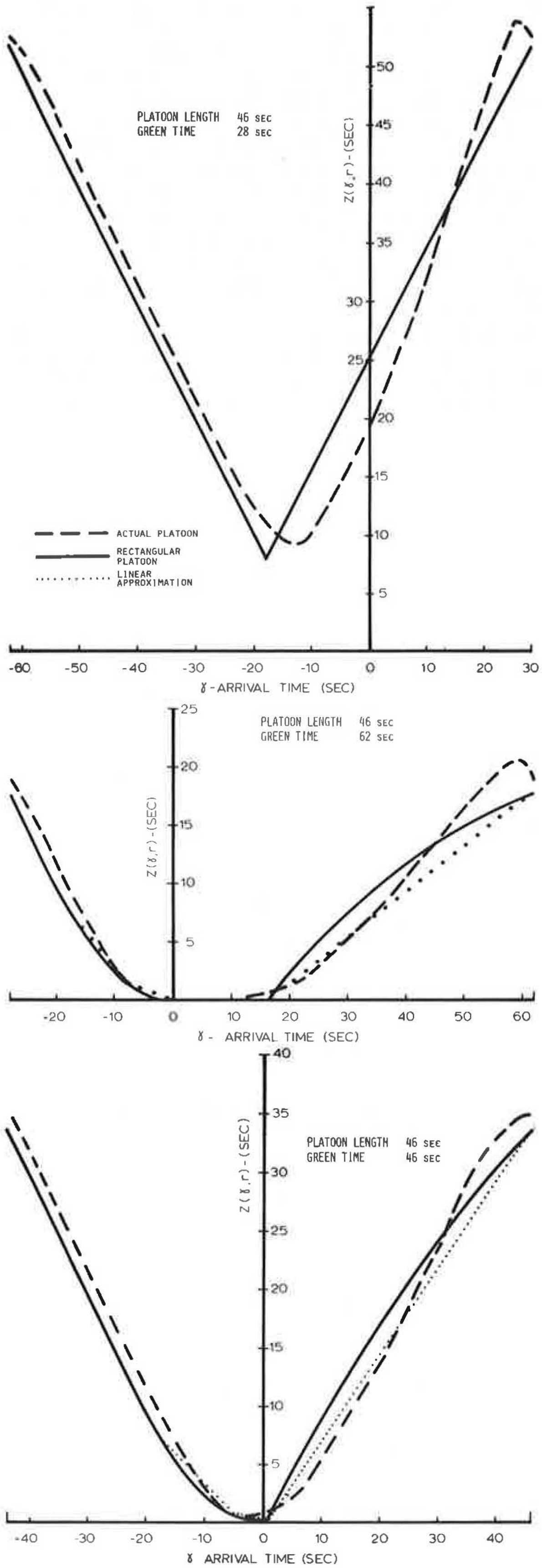


Table 1. Expected overflow queue.

Signal Capacity (vehicles/ cycle)	Degree of Saturation						
	0.20	0.40	0.60	0.80	0.90	0.95	0.975
5	0.00	0.02	0.20	1.15	3.50	8.41	18.36
15	0.00	0.00	0.04	0.70	2.81	7.61	17.50
25		0.00	0.01	0.47	2.41	7.08	16.91
35			0.00	0.34	2.11	6.68	16.45
45				0.23	1.88	6.34	16.05
55					1.68	6.02	15.67

Figure 6. Saturation deterrence function and piecewise linear approximation.

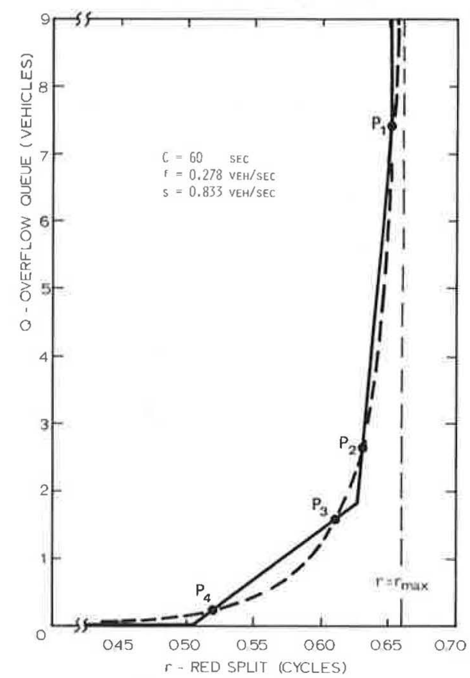
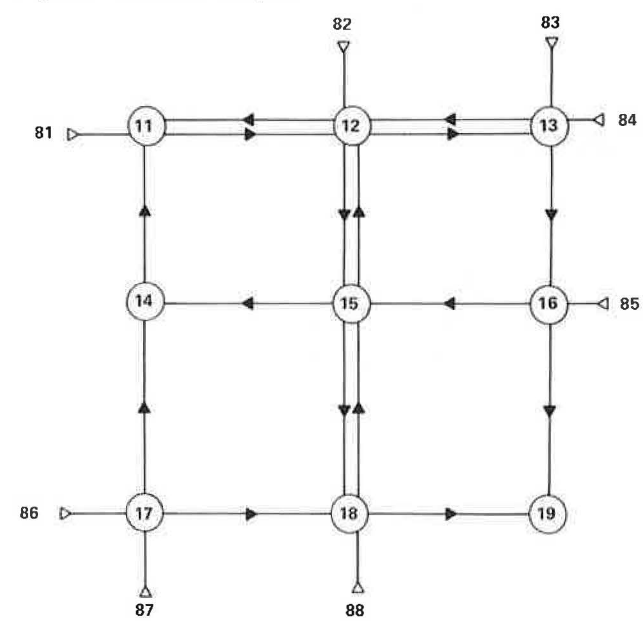


Figure 7. Test network diagram.



has been studied for the signal network example shown in Figure 7. This interplay is shown in Figure 8. It can be seen that the optimal network cycle time is determined as a least cost equilibrium point between delays attributed to the mean traffic flow component and delays attributed to the stochastic component. MITROP determines this point simultaneously with all the other traffic control variables.

The test network was further used to analyze the sensitivity of performance with respect to volume. All volumes in the network were changed by the same percentage, and MITROP was used to optimize offsets and splits for various cycle times. The results are shown in Figure 9. The decisive role played by the cycle time in reaching optimum operating conditions is illustrated here even more emphatically than previously.

CONCLUSIONS

The MITROP program formulates the traffic signal network optimization problem for mixed-integer linear programming. Thus, a global optimal solution to the problem can be achieved via an existing optimization package. Whereas much previous research has been devoted to developing a suitable optimization procedure, the emphasis here is on accurately modeling the traffic flow process and its performance measures. The branch-and-bound procedure used by the MPSX code has proved to be quite efficient in handling the traffic signal network problem because the number of integer variables versus the number of continuous variables in the problem is relatively small. Mixed-integer programming is an active area of research, and further improvements in algorithms can be expected in the future (22). Such developments might make it possible to use MITROP in an on-line mode.

The study also indicates the importance of having all the control variables of the system as simultaneous decision variables. Performance may be significantly degraded when a sequential decision process is used.

Table 2. Input data for test network (Figure 7).

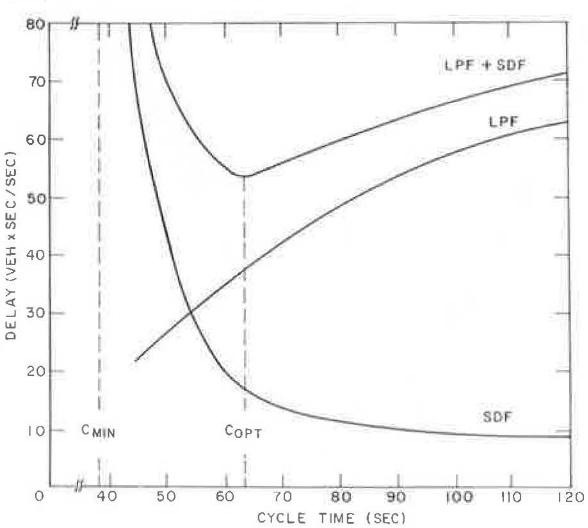
Links	Length (m)	Speed (km/h)	Travel Time (s)	Volume (vehicles/h)	Saturation Flow (vehicles/h)	Platoon Length (cycle)
Internal						
(11, 12)	305	39	28.41	630	1800	0.701
(12, 13)	168	39	15.63	630	1800	0.532
(13, 12)	168	45	13.39	400	1800	0.425
(12, 11)	305	45	24.35	400	1800	0.608
(16, 15)	168	45	13.39	430	1300	0.414
(15, 14)	305	45	24.35	430	1300	0.595
(17, 18)	305	39	28.41	630	1800	0.701
(18, 19)	168	39	15.63	630	1800	0.503
(17, 14)	244	35	24.79	550	3000	0.344
(14, 11)	183	35	18.60	550	3000	0.335
(12, 15)	183	45	14.61	800	3000	0.407
(15, 18)	244	45	19.48	800	3000	0.444
(18, 15)	244	35	24.79	550	3000	0.463
(15, 12)	183	35	18.60	550	3000	0.418
(13, 16)	183	45	14.61	900	2160	0.515
(16, 19)	244	45	19.48	900	2160	0.560
Input						
(81, 11)				630	1800	1.0
(82, 12)				800	3000	1.0
(83, 13)				900	2160	1.0
(84, 13)				400	1800	1.0
(85, 16)				430	1300	1.0
(86, 17)				630	1800	1.0
(87, 17)				550	3000	1.0
(88, 18)				550	3000	1.0

Note: 1 m = 3.3 ft; 1 km/h = 0.6 mph.

Table 3. Loops of the test network and their corresponding integer variables.

Loop	Links and Nodes in the Loop	Optimal Integer
1	(11, 12), (12, 11)	1
2	(15, 12), (12, 15)	1
3	(18, 15), (15, 18)	1
4	(12, 13), (13, 12)	1
5	(14, 11), 11, (11, 12), 12, (12, 15), 15, (15, 14), 14	-1
6	(15, 12), 12, (12, 13), 13, (13, 16), 16, (16, 15), 15	-1
7	(17, 18), 18, (18, 15), 15, (15, 14), 14, (17, 14), 17	0
8	(16, 15), 15, (15, 18), 18, (18, 19), 19, (16, 19), 16	0

Figure 8. Variation of network performance function with cycle time.



Of particular importance is the cycle time, which is strongly affected by the flows in the network. In this context, it is essential to adequately model the stochastic behavior of traffic, which in MITROP is represented by SDF. If, for instance, the cycle time for the test network were determined by Webster's method, the result would be close to 80 s (node 13 in Figure 7 is the most heavily loaded intersection). Inspection of Figure 8 shows that the objective function for this cycle time is roughly 12 percent higher than the optimum value at 63.8 s.

A useful extension of MITROP is that additional performance measures can be assigned to sections of the network, if so desired. MITROP assumes that the LPF is a function of offsets only on that link, similar to assumptions made by SIGOP and the combination method (23). It has been shown that in certain cases, particularly on lightly traveled links, it may be advantageous to follow vehicular movement on two links or more (4, 17). This can be easily done in MITROP by maximizing bandwidth on selected arterial sections, which provides this coupling feature, while simultaneously minimizing delay on other sections of the network. The bandwidth maximization problem was formulated in the past in terms of mixed-integer linear programming by Little (24) and is therefore compatible with the MITROP optimization model.

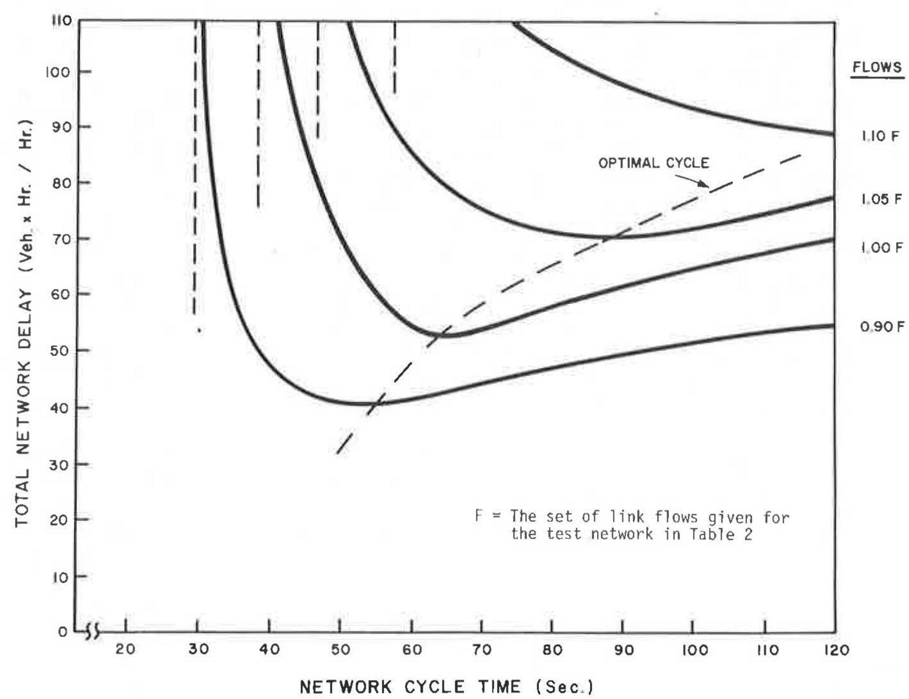
In addition to its use in the network example given in this paper, MITROP was also used for a portion of the urban traffic control system/bus priority system (UTCS/BPS) in Washington, D.C., containing 20 nodes, 63 links, and 21 independent loops, and for an arterial

Table 4. Optimal settings and output data for test network.

	Offset		Green Time				
Links	Seconds	Cycle	Seconds	Cycle	LPF (vehicle-h/h)	SDF (vehicles)	Degree of Saturation
Internal							
(11, 12)	44.6	0.70	28.7	0.45	1.416	1.255	0.778
(12, 13)	33.8	0.53	24.9	0.39	2.175	2.244	0.897
(13, 12)	30.0	0.47	28.7	0.45	0.543	0.0	0.489
(12, 11)	19.1	0.30	37.6	0.59	0.315	0.0	0.373
(16, 15)	30.6	0.48	33.1	0.52	1.752	0.0	0.641
(15, 14)	33.8	0.53	37.6	0.59	0.566	0.0	0.565
(17, 18)	24.9	0.39	30.6	0.48	1.795	0.690	0.729
(18, 19)	24.2	0.38	24.9	0.39	0.724	2.244	0.897
(17, 14)	29.3	0.46	17.2	0.27	0.339	0.0	0.674
(14, 11)	28.0	0.44	17.2	0.27	0.736	0.0	0.674
(12, 15)	25.5	0.40	21.7	0.34	1.318	0.943	0.775
(15, 18)	31.2	0.49	24.2	0.38	1.263	0.0	0.693
(18, 15)	32.5	0.51	21.7	0.34	0.600	0.0	0.535
(15, 12)	38.2	0.60	24.2	0.38	1.889	0.0	0.479
(13, 16)	21.0	0.33	30.6	0.48	0.780	1.660	0.868
(16, 19)	25.5	0.40	30.0	0.47	0.750	1.930	0.887
Input							
(81, 11)			37.6	0.59	1.650	0.0	0.593
(82, 12)			24.2	0.38	3.771	0.0	0.693
(83, 13)			30.0	0.47	3.855	1.930	0.887
(84, 13)			24.9	0.39	1.767	0.0	0.564
(85, 16)			23.6	0.37	2.195	2.714	0.901
(86, 17)			33.1	0.52	2.119	0.0	0.673
(87, 17)			21.7	0.34	2.742	0.0	0.535
(88, 18)			24.2	0.38	2.450	0.0	0.479

Note: Cycle time = 63.8 s; objective function = 53.120 vehicle-h/h.

Figure 9. Sensitivity of network performance function with respect to cycle time and flows.



street with 11 signals in Waltham, Massachusetts. The results are described in detail in a technical report (13). The MITROP computer program is currently operational and available from the Operations Research Center at the Massachusetts Institute of Technology.

ACKNOWLEDGMENT

The research reported in this paper was supported in part by the Federal Highway Administration, U.S. Department of Transportation, and in part by the Army Research Office. The contents of this paper reflect the views of the authors and do not necessarily reflect the official views or policies of the sponsoring agencies.

REFERENCES

1. SIGOP: Traffic Signal Optimization Program, A Computer Program to Calculate Optimum Coordination in a Grid Network of Synchronized Traffic Signals. Traffic Research Corp., New York, 1966.
2. J. A. Kaplan and L. D. Powers. Results of SIGOP-TRANSYT Comparison Studies. Traffic Engineering, Vol. 43, No. 12, 1973.
3. P. D. Whiting. The Glasgow Traffic Control Experiment: Interim Report on SIGOP and TRANSYT. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, Rept. LR 430, 1972.
4. C. J. MacGowan and H. S. Lum. SIGOP or TRANSYT? Traffic Engineering, Vol. 45, No. 4, 1975, pp. 46-50.
5. J. Holroyd. The Practical Implementation of Combination Method and TRANSYT Programs. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, Rept. LR 518, 1972.
6. F. V. Webster. Traffic Signal Settings. Road Research Technical Paper 30, Her Majesty's Stationery Office, London, 1958.
7. N. Gartner and J. D. C. Little. Generalized Combination Method for Area Traffic Control. TRB, Transportation Research Record 531, 1975, pp. 58-69.
8. D. I. Robertson. TRANSYT: A Traffic Network Study Tool. U. K. Road Research Laboratory, Crowthorne, Berkshire, England, Rept. LR 253, 1969.
9. F. A. Wagner, F. C. Barnes, and D. L. Gerlough. Improved Criteria for Traffic Signal Systems in Urban Networks. NCHRP, Rept. 124, 1971.
10. L. Rach, J. K. Lam, D. C. Kaufman, and D. B. Richardson. Evaluation of Off-Line Traffic Signal Optimization Techniques. TRB, Transportation Research Record 538, 1975, pp. 48-58.
11. P. J. Tarnoff. The Results of FHWA Urban Traffic Control Research: An Interim Report. Traffic Engineering, Vol. 45, No. 4, 1975, pp. 27-35.
12. Mathematical Programming System Extended—Mixed Integer Programming (MIP) Program Description. Technical Publications Department, IBM Corp., White Plains, N.Y., Program 5734-XM4, 1971.
13. N. Gartner, J. D. C. Little, and H. Gabbay. Optimization of Traffic Signal Settings in Networks by Mixed-Integer Linear Programming. Operations Research Center, MIT, Cambridge, Technical Rept. 91, March 1974.
14. N. Gartner. Constraining Relations Among Offsets in Synchronized Signal Networks. Transportation Science, Vol. 6, 1972, pp. 88-93.
15. K. W. Huddart. The Importance of Stops in Traffic Signal Progressions. Transportation Research, Vol. 3, 1969, pp. 143-150.
16. C. C. Chung and N. Gartner. Acceleration Noise as a Measure of Effectiveness in the Operation of Traffic Control Systems. Operations Research Center, MIT, Cambridge, Working Paper OR 015-73, 1973.
17. K. W. Huddart and E. D. Turner. Traffic Signal Progressions—GLC Combination Method. Traffic Engineering and Control, Vol. 11, 1969, pp. 320-327.
18. J. A. Hillier and R. Rothery. The Synchronization of Traffic Signals for Minimum Delay. Transportation Science, Vol. 1, 1967, pp. 81-94.
19. W. Jacobs. A Study of the Averaged Platoon Method for Calculating the Delay Function at Signalized Intersections. Department of Civil Engineering, Univ. of Toronto, BASc thesis, 1972.
20. N. Gartner. Microscopic Analysis of Traffic Flow Patterns for Minimizing Delay on Signal-Controlled Links. HRB, Highway Research Record 445, 1973, pp. 12-23.
21. R. Wormleighton. Queues at a Fixed Time Traffic Signal With Periodic Random Input. Journal of Canadian Operations Research Society, Vol. 3, 1965, pp. 129-141.
22. W. D. Northup and J. F. Shapiro. Design and Computational Experience With an Interactive Mixed Integer Programming System. Paper presented at ORSA-TIMS Joint National Meeting, Chicago, April-May 1975.
23. J. A. Hillier. Appendix to Glasgow's Experiment in Area Traffic Control. Traffic Engineering and Control, Vol. 7, No. 9, 1966, pp. 569-571.
24. J. D. C. Little. The Synchronization of Traffic Signals by Mixed-Integer Linear Programming. Operations Research, Vol. 14, 1966, pp. 568-594.

Discussion

Dale W. Ross, DARO Associates, Inc.

The new computer method MITROP, which the authors have developed for optimizing traffic signal timings, is a significant contribution to the field of traffic control. It is the first comprehensive, mathematically rigorous method to be developed for the simultaneous optimization of cycle time, splits, and offsets in an arbitrary traffic signal network.

Although the name of the method, mixed-integer linear programming, is descriptive of the method used, it does not spotlight the essential contribution of the authors. It is not the use of mixed-integer linear programming that is new; Little (24) applied mixed-integer linear programming 10 years ago. Rather, it is the modeling of the traffic flow and of the signal timing relations that is new. The authors have done a commendable job of (a) realistically modeling traffic flow, (b) accounting for the important real-life constraints on signal timings in their model, (c) leaving all variables—cycle time, splits, and offsets—as optimizable variables in their model, and (d) judiciously approximating the link delay functions by piecewise linear convex functions so that the model is amenable to optimization by using a standard, off-the-shelf mathematical programming package that guarantees a global optimum.

The authors' modeling has some important ramifications. The treatment of random variations in traffic flow through a saturation deterrence function (SDF) for each link probably has the most effect. The authors have shown how essential it is to include the SDF so that capacity loss at short cycles is prevented. Indeed, it is the interplay of the deterministic and stochastic delay components that makes cycle length optimization important, as the authors have illustrated in Figures 8 and 9. The inclusion of the SDF model also highlights the importance of allowing signal splits to be freely optimized. Previous signal timing programs such as SIGOP and TRANSYT are deficient in freely optimizing cycle time and splits. SIGOP does not model the stochastic delay component at all and only evaluates cycle times that are preselected by the user. As the authors point out, in past applications, SIGOP consistently selected the lower bound on cycle length. TRANSYT uses a critical intersection approach for determining network cycle time and ignores the interaction between traffic flows over the spatial structure of the traffic signal network.

If there is one aspect of the MITROP modeling that could be improved it is the platoon modeling. Currently, platoons are assumed by MITROP to be uniform in flow rate over the green time of an upstream signal. This has a decoupling effect of making platoon flow along one link independent of that along adjoining links. This may affect the ability of the model to develop good progressive movements along major streets within a grid network. Another possible model improvement would be to allow the platoon speed along individual links to be variable within limits; in fact, platoon speed could possibly be yet another optimization variable. Little (24) allowed speed to be variable, and Leuthardt (25) has recently shown that allowing speed to be optimizable can lead to improved signal progressions.

The required computer time for MITROP appears to be less than that for TRANSYT and more than that for SIGOP for the same or similar networks. I draw this tentative conclusion from (a) computer time data given by the authors (13) for an IBM 370/165 computer, (b) computer time formulas given for TRANSYT and SIGOP for an IBM 360/65 (4), and (c) a typical ratio of 3.15 for the computer time of similar scientific programs on an IBM 360/65 versus a comparable IBM 370/165 system (27). Although the authors do not give computer storage requirements for MITROP, its use of the IBM mathematical programming package probably makes it more consumptive of computer memory than either SIGOP or TRANSYT. Its storage requirements probably preclude its use, in present form, for real-time traffic control. Because few people have used MITROP, it remains to be seen how easy it is to use. Specifying input data may be a formidable effort since a large number of constraints must be defined for the mixed-integer linear program. However, it is my understanding that the authors have developed a preprocessor to assist in the data input.

I have some suggestions for improving the mathematical optimization portion of MITROP.

1. One way of reducing the computer memory (particularly main memory, as opposed to bulk memory) requirements of MITROP and thereby making it more amenable to real-time application might be to scan cycle times in an allowable range and to use the Dantzig-Wolfe decomposition principle (27) to solve the resultant (mixed-integer) linear programs. When cycle time is fixed, at one point in the scan, the MITROP constraints become largely uncoupled, and the constraints on the sum of offsets around closed loops provide the main coupling. Each linear program can then be broken into subprograms. In each subprogram, only a small number of the constraint columns need to be examined in main memory of the computer; the rest can be kept in bulk memory.

2. In the branch-and-bound method used to solve the mixed-integer linear program, it might be advantageous to solve the dual linear program to establish bounds. In the dual problem, the integer variables would appear in the linear objective function rather than in the constraints. In this case, when the integers are changed in the branch-and-bound process, feasible solutions of a linear program corresponding to one set of integers are also feasible solutions of a linear program with a different set of integers. This would mean, it is hoped, that fewer linear program iterations would be needed to find a new (dual problem) solution for a change of integers. Because the value of the optimal primal linear program is the same as that of the optimal dual linear program, the dual can be used to establish the bounds.

In summary, the fact that MITROP (a) allows simultaneous optimization of cycle time, splits, and offsets, (b) ensures a global optimum set of signal timings, and (c) has computer requirements comparable to other currently used signal timing methods makes it, in my opinion, a superior method for (off-line) computation of signal timings. Future use of MITROP in a variety of applications will lead to improved ease of use and to an improved understanding of its sensitivity to real data in actual traffic networks.

REFERENCES

25. H. R. Leuthardt. NO-STOP-1 Versus SIGPROG. TRB, Transportation Research Record 531, 1975, pp. 70-81.
26. R. F. Littrell. Economics of Scale in the IBM 360 and 370. Datamation, March 1974, pp. 83-88.
27. G. B. Dantzig and P. Wolfe. Decomposition Principle for Linear Programs. Operations Research, Vol. 8, No. 1, 1960.

Authors' Closure

We thank Ross for his thoughful discussion of our paper. We have further comments on two of the points that he raised.

With respect to platoon modeling, our simple formulation has worked well on the Hillier and Rothery data (18), and so we adopted it. We recognize that the model deemphasizes progression, and we have a number of approaches that would introduce more progression. One of these is particularly straightforward to implement. All the variables required to calculate bandwidth on any street are already in the computer. It would be rather easy to calculate progression bands for key arterials and introduce them into the objective function. The mixed-integer algorithm could then minimize a weighted combination of bandwidths and delays, thereby introducing progression wherever it is not dominated by delay considerations.

With respect to optimization efficiency Ross makes some worthwhile suggestions. We gave our primary research priority to the model, its fidelity, and the method of its translation into an optimization framework. We left the actual mathematical programming algorithm to standard packages. It would be interesting to explore special-purpose algorithms for the problem, although developments in general-purpose algorithms and increases in computing efficiency that resulted from hardware advances have somewhat decreased the pressure for special-purpose programs.