Compromise Approach To Optimize Traffic Signal Coordination Problems During Unsaturated Conditions

CHANG-JEN LAN, CARROLL J. MESSER, NADEEM A. CHAUDHARY, AND EDMOND CHIN-PING CHANG

A methodology to optimize the traffic signal coordination problem on an arterial network simultaneously considering delay minimization and progression bandwidth maximization criteria is presented. This approach generates a compromise solution to these two conflicting criteria and, in some cases, produces timing solutions with less delay and less bandwidth than the conventional MAXBAND solutions, although sometimes the outcome is reversed. In general, there is usually a trade-off between delay and bandwidth under the well-timed traffic signal system.

Two conventional approaches are available for coordinating traffic signals in an urban network: delay minimization and progression bandwidth maximization. Timing plans based on the bandwidth approach are preferred by drivers because of the ease in which progression bandwidth can be visualized. A traffic engineer also prefers the bandwidth-based approach because it provides dependable solutions. However, delay-based timing plans may produce better overall system performance and may be preferable from the perspective of traffic systems management. Therefore, traffic engineers may desire to optimize both signal timing objectives based on the conflicting nature of the delay and progression bandwidth criteria (*1*). Therefore, usually a trade-off is necessary.

Cohen and Liu (*2*) proposed a bandwidth-constrained delay minimization methodology that uses TRANSYT-7F to fine-tune offset and green split while preserving the progression bands generated from the MAXBAND program (*3*). This approach may improve system performance as compared with the centered bandwidth timing plans. However, in some situations preserving the progression bands only produces local optimal solutions from a systemwide viewpoint, since there is a trade-off between bandwidth and delay (see Figure 1). This diagram is conceptually constructed on the basis of an investigation of simulation results of real timing data collected from several study sites. The solution with least delay and best bandwidth is an ideal optimal solution used as a benchmark for evaluating the system performance of signal timing plans. Within the timing solution space (shaded area), the solutions along the frontier line are of major interest. Solutions S_1 through S_3 are bandwidth target solutions that carry the maximal bandwidth and relatively high delay. Solutions S_8 through S_{10} are solutions with lower bandwidth but the least delay. These timing solutions can be generated by existing technologies. In the bottom right corner, the general fashion of frontier line bypasses the ideal point. Solutions S_4 through S_7 follow a trend in which delay is reduced as progression decreases, depending on how the decision maker trades off O_1 against O_2. The evaluation of these timing solutions also depends on the preference of the decision maker. Therefore, it is suggested that reallocating the green time resource on the basis of the trade-off between lost bandwidth and delay savings would be beneficial to further reducing system delay and upgrading system performance.

A formal method is proposed to optimize arterial signal timing using both delay minimization and throughput maximization criteria. The enhanced arterial model called COMBAND follows the basic MAXBAND (*3*) formulation and concepts described in the original MITROP model (*4,5*). The results show that a trade-off between delay and progression objectives can render better signal timing plans while maintaining the sound features of both.

MODEL DESCRIPTION

The COMBAND model improves bandwidth-based timing plans by combining both bandwidth and delay/stops considerations. In the MAXBAND model, only the bandwidth decision variable is set up in the objective function, which implies that bandwidth maximization is the only criterion being considered. In most situations, such a single-criterion formulation may generate ineffective solutions because the sidestreet delay increases considerably when its traffic flow

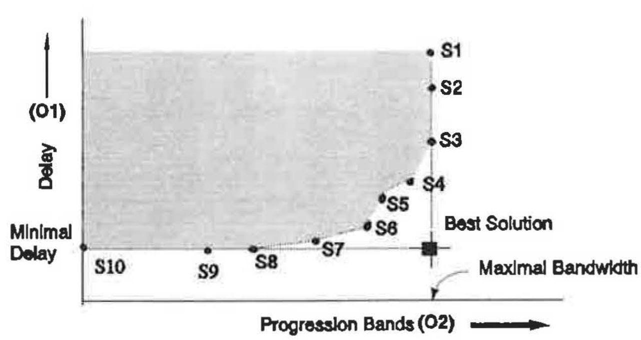

FIGURE 1 Signal timing solutions based on two criteria.

Texas Transportation Institute, Texas A&M University, College Station, Tex. 77843.

increases. Some modifications were made to extend the capability of bandwidth-based methodology dealing with delay minimization, such as the directional weighting of the inbound and outbound bandwidth and multiweight bandwidth for each directional road link of the arterial (*6,7*); however, in terms of both timing objectives a global optimal solution still cannot be achieved.

The traffic signal optimization problem contains many variables that in turn affect the system performance, individually or collectively. These variables include delay, stops, progression bands, capacity (or throughput), fuel consumption, emissions, and journey time. Delay, stops, progression bands, and capacity have been analyzed often in previous research in that these variables are not strongly interrelated but sensitively responsive to traffic systems in most situations. Fuel consumption and emissions are normally modeled as secondary functions of delay and stops, hence, they are highly correlated with delay and stops as dependent variables to system performance. They need not be considered independently. As for journey time, it could be affected by several combinations of variables, rendering it difficult to be modeled. In this paper, delay/stops and throughput are chosen as the criteria to evaluate signal timing plans.

Essentially, delay/stops and progression bands become two conflicting criteria in well-timed systems. The defined problem is formulated as a multicriteria decision making (MCDM) problem and solved by linear multiobjective programming techniques. Another approach is to formulate the problem as linear programming (LP) formulation, combine the operational criteria in the LP objective function, and solve the problem using LINDO, MPSX/370E, or other LP optimization packages. Both approaches have to convert the nonlinear objective function into a linear form before applying the LP technique to solve the problem since there is no tool available to solve the mixed-integer nonlinear programming problem directly. Before proceeding further, several assumptions should be made. The following assumptions are also applied in the underlying model:

1. Prevailing traffic conditions are not saturated;
2. Traffic arrival flow rate and service rate remain constant. The time-stationary flow assumptions will be relaxed in the future work, with provisions of overflow queue or temporary oversaturation condition;
3. No platoon dispersion occurs on the coordinated arterial; and
4. No midblock flow occurs.

The COMBAND model uses notation similar to that of the MAXBAND model to promote ease of reference to existing technology. The complete model formulation and notation are provided in the Appendix. Figure 2 shows the notation in a time-space diagram.

Objective Function

Delay/Stops Criteria

From the preceding assumptions, we postulate that the queue accumulated during red dissipates during green and no overflow queue occurs. Thus, the total delay can be written without the overflow term and is defined as follows:

$$TD = \frac{qR^2}{2C(1 - q/s)} = \frac{qCr^2}{2(1 - q/s)} \text{ (veh·sec/sec)}$$

where

q = arrival rate (vps),
R = effective red time (sec),
s = saturation flow rate (vps),
C = cycle time (sec), and
$r(g)$ = red (green) time in fraction of cycle.

Without considering overflow delay, the total delay is observed to be a quadratic function of red time as shown in Figure 3. The delay curve versus red time to be linearized is bounded by r_{min} and r_{max} and any amount of red time can be allocated within this range depending on the "steepness" of the delay curve. The first task is to determine how many components should be used to fit the curve accurately. Yagar indicated that three piecewise-linear components can fit the delay curve well enough to produce quite accurate results (*8*). Gartner et al. also split the curve into three piecewise linear

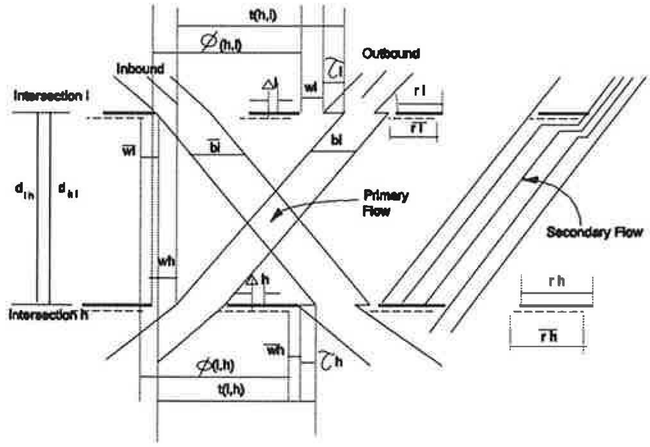

FIGURE 2 Time-space diagram for MAXBAND and COMBAND models.

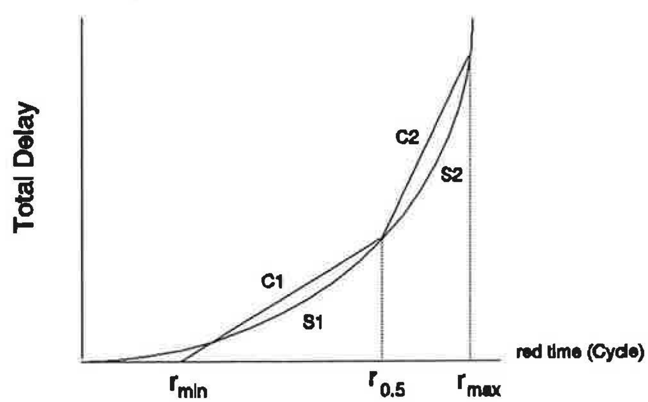

FIGURE 3 Total delay versus red time.

components. However, we have elected to apply only two piecewise linear components for the following reasons:

1. Under low to moderate flow conditions, the q/c value seldom exceeds 0.7. The delay curve is not too steep when (q/c) is below this level.
2. From the first reason, the two-component linearized approximation generates quite similar results to three-component linearized approximation.
3. The decision variables in the LP problem are reduced to enhance the computational efficiency, especially when dealing with a large-scale network problem.

Another question might be raised as to where the cutting line between two linear pieces lies. For each link, we have

$$x = \frac{q}{c} = \frac{q}{s \cdot g} = \frac{q}{s(1 - r)}$$

$$r = 1 - \frac{q}{sx}$$

Here $x = 0.5$ is selected as the cutting line after accomplishing several experiments; the associated r-value $(r_{0.5})$ is also calculated. Thus, the slope of each linear component can be determined as follows, respectively:

For C_1

$$\frac{Ar_{0.5}^2 - Ar_{\min}^2}{r_{0.5} - r_{\min}} = A(r_{0.5} + r_{\min})$$

For C_2

$$A(r_{0.5} + r_{\max})$$

where A equals $q/[2(1 - q/s)]$. On the other hand, it may be desirable to take into account stops along with delay in the disutility function. The number of vehicle stops per cycle is

$$\frac{qR}{(1 - q/s)} = \frac{qrC}{(1 - q/s)}$$

Let B be $q/(1 - q/s)$, multiplied by a factor $k(k = 0 \text{ to } 1)$; added with $(1 - k) \cdot C_1$ and $(1 - k) \cdot C_2$, respectively, the disutility function becomes a weighted combination of delay and stops with the following coefficients:

$$D1 = (1 - k) \cdot C1 + k \cdot B$$

$$D2 = (1 - k) \cdot C2 + k \cdot B$$

The factor k is specified by the user. If k is set to zero, the disutility objective function becomes minimization of systemwide delay without considering stops. The authors suggest that k be chosen in the range from 0.3 to 0.7. So far, for intersection i, we have two decision variables, $r1_{mi}$ and $r2_{mi}$, with coefficients $D1_i$ and $D2_i$ respectively in the objective function, that is (notation is only for outbound direction):

Minimize

$$\sum_i (D1_i r1_{mi} + D2_i r2_{mi})$$

where $r1_m$ and $r2_m$ are subjected to the constraints $r_{\min} \leq r1_m \leq r_{0.5}$ and $0 \leq r2_m \leq (r_{\max} - r_{0.5})$, respectively. Moreover, we adjust the arterial delay function value by the progression factor (PF) as used in the *Highway Capacity Manual* (*9*) to account for the effect of platoon progression on delay, assuming there exists a moderately favorable platoon condition under COMBAND-generated timing plans.

Throughput Criteria

In the COMBAND model, an attempt is made to maximize the arterial throughput and minimize the system total delay/stops to increase the system throughput. For intersection i, the arterial throughput is essentially equivalent to the sum of the proportions of vehicles (including arterial through traffic q_1 and sidestreet turning traffic q_2) from upstream intersection passing through the downstream intersection inside and outside the bandwidth during green; for example:

$$q_1 b_i + q_2(g_i - b_i)$$

$$q_1 b_i + q_2(1 - r1_{mi} - r2_{mi} - b_i)$$

$$(q_1 - q_2)b_i - q_2(r1_{mi} + r2_{mi}) + q_2$$

where q_2 can be removed from objective function. Finally, combining delay/stops and throughput objectives, the objective function becomes (notation is only for outbound direction)

Maximize

$$\sum_i [(q_1 - q_2)b_i - q_2(r1_m + r2_m) - (D1_i r1_{mi} + D2_i r2_{mi})]$$

Constraints

In addition to the basic MAXBAND formulation, a number of the following constraints are added to enhance the capability of the underlying model. First, the technique of releasing green splits as variables is used to make the optimization of green split possible (*10*). A set of constraints and binary integer variables are included. As shown in Figure 3, total delay increases considerably as traffic flow approaches capacity (i.e., the volume-capacity ratio (q/c) is approaching 1.0). To avoid the escalating delay produced by an overflow queue and maintain a minimum level of service, we restrict volume/capacity ratio below 0.95 and include associated red time upper-bound constraints as used in the MITROP program; for example:

$$r \leq r_{\max} = 1 - \frac{q}{0.95s}$$

On the other hand, it is desired to set the minimum effective green time on each approach according to the following considerations:

1. The feasible minimum amount of effective green time pedestrians need to traverse the side streets safely and the time would satisfy the driver's expectancy.
2. Nominal green splits calculated by Webster's method under the control of the local minimum-delay cycle.

An important issue to be considered pertains to the allocation of the slack green time, defined as the excess time of the system optimal cycle beyond the local minimal-delay cycle. Let z and z_l stand for the reciprocals of the system optimal cycle and the local minimal-delay cycle, and zr represents the ratio of these two variables. For each intersection

$$zr = \min\{1, z/z_l\}$$

This equation is equivalent to the following constraints

$$zr \leq 1$$

$$zr \leq z/z_l$$

$$1 - zr \leq M \cdot \delta$$

$$z/z_l - zr \leq M \cdot (1 - \delta)$$

where M is a big number and δ is the binary integer variable. Because z is in unit of cycle time, $M = 1$ is big enough. Therefore, after some rearrangement, the above constraints become

$$zr \leq 1$$

$$zr \leq z/z_l$$

$$zr + \delta \geq 1$$

$$zr - z/z_l - \delta \geq -1$$

As with the intersections carrying the slack time, the green splits of main and side streets will be lowered in proportion to zr, and the slack green time is further reallocated by the underlying model instead of Webster principle.

To make the model closer to real situations, it is assumed that the journey time increases with increasing traffic flow. The BPR travel time prediction function sometimes used in transportation planning models, $t_i = t_0[1.0 + 0.15(q/c)^4]$, is used to characterize the link volume-delay relationship, where t_i is the predicted travel time on link i given a specific traffic flow, t_0 is the free-flow travel time, and c is link capacity. Here in BPR function, the term (q/c) is determined by $2Y/(1 + Y)$, (Y being the sum of flow ratios of the critical movements) under the assumption that the degree of saturation is the same for all critical phases of the intersection for optimum division of the cycle (*11*).

As with the range of system cycle concerned, the upper and lower bounds are selected according to the following considerations. The concept is illustrated in Figure 4.

1. Define the intersection with largest local minimal-delay cycle C_0, as the critical intersection. The calculation of minimal-delay cycle is based on Webster's method, for example, $C_0 = (5 + 1.5L)/(1 - Y)$, where L is total lost time per cycle (*11*).

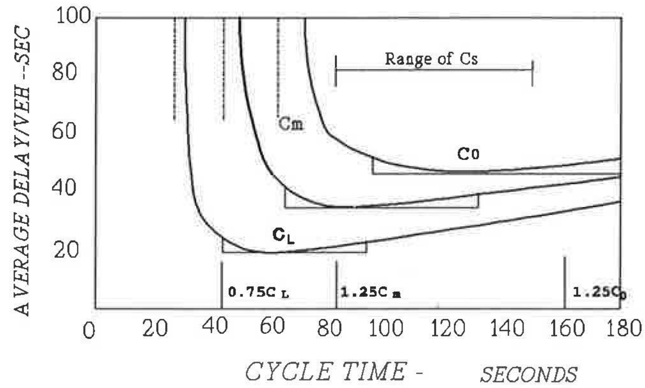

FIGURE 4 Effect on delay of variation of cycle length.

2. According to Webster's method, the minimum cycle is just long enough to allow all the traffic arriving during a cycle to optimally clear the intersection. For deterministic flows, the cycle is given by $C_m = L/(1 - Y)$ (the vertical asymptote to the delay-cycle curve), at which the degree of saturation is close to 1. To ensure the critical intersection is operated under capacity where the level of flow varies appreciably, the system was set at minimum cycle, C_s, at least equal to $1.25L/(1 - Y)$. The lower bound of system cycle is also confined by $0.75*$ (smallest local minimal-delay cycle, C_L).
3. As indicated (*11*), the delay for cycle within the range 0.75 to 1.5 of the optimal value is never 10 to 20 percent more than minimum delay. We further restrict the upper limit as $1.25C_0$ to avoid too much waste of green time.
4. From a practical standpoint, the cycle should be within the range of 40 to 150 sec.

In summary, we suggest the range of system cycle as follows:

$$\max\{40, 0.75C_L, 1.25C_m\} \leq C_s \leq \min\{1.25C_0, 150\}$$

In the basic MAXBAND model, queue clearance time must be supplied by users. However, under the assumption of uniform arrivals, queue length and queue clearance time can be estimated approximately. Assuming that the primary flow can fully utilize the bandwidth without being stopped under a favorable platoon condition, the queue would be produced mainly by the secondary flow consisting of the turning movements of upstream intersection from side streets during arterial red time and through movement from upstream intersection during slack time as shown in Figure 5. Let q_L, q_R, and q_T be the sidestreet left-turn, right-turn, and arterial through movement traffic coming from upstream intersection h, then the number of vehicles traveling on the link between intersection h and i during a cycle are produced by

- Left-turn movement from inbound side street:

$$q_L \cdot \bar{l}_{ch}$$

- Right-turn movement from outbound side street:

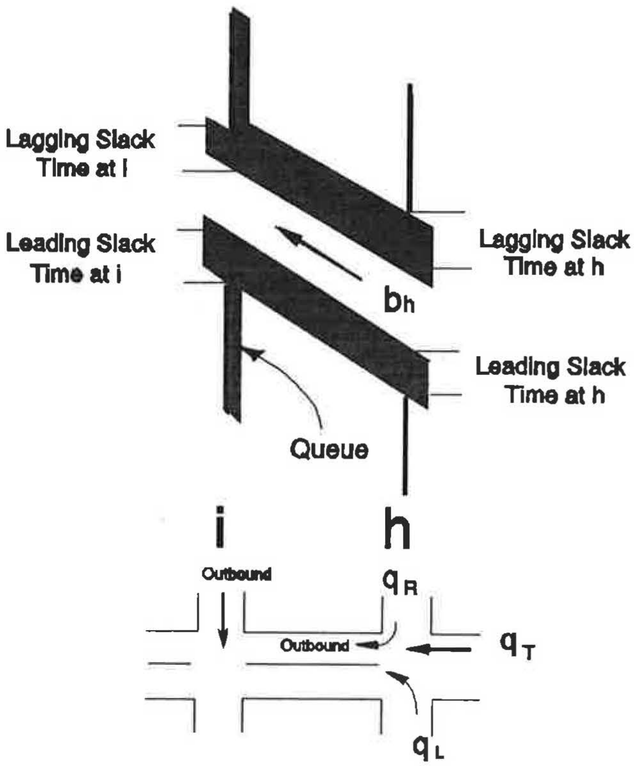

FIGURE 5 Queue produced by secondary flow.

Assuming that every intersection is on the right turn on red (RTOR) operation for right-turn movement, outbound sidestreet right-turn movement is allowed only during outbound sidestreet through phase and inbound arterial left-turn phase. Therefore, the right-turn traffic is equal to

$$q_R \cdot (1 - r1_{ch} - r2_{ch} + \bar{l}_{ch})$$

- Through movement on arterial during leading slack time (w_h) to the progression band:

$$q_T \cdot w_h$$

- Through movement on arterial during lagging slack time $(1 - r1_{mh} - r2_{mh} - w_h - b_h)$ of previous cycle:

$$\max[q_T \cdot (r1_{mi} + r2_{mi} - r1_{mh} - r2_{mh} + w_i - w_h), 0]$$

The flow generated by these four categories is denoted as Q_i and the average discharge headway on the link i as h_i. The required queue clearance time is equal to the start-up loss time plus $h_i \cdot Q_i$, where the start-up time is assumed to be 2 sec. If the available slack time w_i is greater than the required queue clearance time, then there will be no queueing. Thus the queue clearance time is

$$\tau_i = \max[que, 0]$$

where $que = 2.0 \cdot z + h_i \cdot Q_i - w_i$. The equation is equivalent to the following constraints:

$$\tau_i \geq 0$$

$$\tau_i - que \geq 0$$

$$\tau_i - \delta_i \leq 0$$

$$\tau_i - que + \delta_i \leq 1$$

Here, constraints $\tau_i \geq 0$ are implicitly processed in normal LP methodology and need not be specified. The queue forming during lagging slack time can be solved in a similar fashion.

MODEL TESTING AND RESULTS

The LINDO optimization package was used to calculate signal timing plans for the MAXBAND, COMBAND models, and bandwidth-target solution (*12*). Also included were TRANSYT optimization solutions with and without bandwidth constraints and PASSERII-90 solutions into analysis. To measure the effectiveness of performance on the common basis, the TRAFNETSIM network simulation package was employed to evaluate these timing solutions. The exogenous data including traffic volumes, lane configuration, and such were collected from three arterial networks: Skillman Avenue (with four intersections), 12th Street (a.m. peak), and 12th Street (p.m. peak; with seven intersections).

For each timing solution, at least 5 simulation runs with different random number seeds were performed, and each run took 15-min. We choose 15-min periods because the traffic arrival pattern starts to become unstable in a longer time frame so that the initial time solution loses the capability to accommodate the forthcoming traffic conditions. Besides, the TRAF-NETSIM simulation model reflects some degree of variation in the simulation results. Several replications are needed to reduce the variation and to suggest reliability of mean value. Here, the required sample size is five replications, providing a limit on a 95 percent probability that the sample mean will be within a range of acceptable error.

To make consistent comparisons among different methods, cycle length was held constant for each case. The other three timing variables (offsets, green splits, and phasing sequences) were optimized by the MAXBAND and COMBAND models. Moreover, the nonuniform bandwidth concept is used in the COMBAND model since the link-specific bandwidth is weighted by traffic volume (7). The total delay function described previously implies that the red split variables are also weighted by traffic volume. Therefore, the decision variables, bandwidth and red splits, are weighted with respect to their contributions to the overall objective function on the same scale.

The results documented in Tables 1, 2, and 3 are briefly described in the next paragraphs. Figures 6, 7, and 8 show the relationship between delay and bandwidth for each case. These figures also confirm the concept depicted in Figure 1.

1. In some cases, the COMBAND model produces timing solutions with less delay and less bandwidth than the conventional MAXBAND solutions and sometimes the outcome is reversed. However, it is shown, that there is a trend of

TABLE 1 Solutions Based on Bandwidth

Case Name	Model Name	Efficiency (%)	Through-put(vph)	Attain-ability(%)	Selected Cycle Length
SKILLMAN AVENUE	TRANSYT[1]	0.1580	18820	36.6	95 Seconds
	BC-TRANSYT[2]	0.4291	18824	100	
	MAXBAND	0.4291	18804	100	
	COMBAND	0.4340	18944	100	
	PASSER II-90	0.3947	18252	100	
12TH STREET (AM PEAK)	TRANSYT	0.1500	17996	65.9	90 Seconds
	BC-TRANSYT	0.2450	17500	100	
	MAXBAND	0.2450	16440	100	
	COMBAND	0.2650	15560	100	
	PASSER II-90	0.1889	15596	76.0	
12TH STREET (PM PEAK)	TRANSYT	0.1164	21020	38.6	116 Seconds
	BC-TRANSYT	0.2677	21224	88.6	
	MAXBAND	0.2677	21020	88.6	
	COMBAND	0.2578	20539	95.7	
	PASSER II-90	0.2328	20988	80.0	

Note: 1. The bandwidth values in the calculation of efficiency and attainability are directly read from time-space diagram produced by TRANSYT.
2. Bandwidth-Constrained TRANSYT solutions which take MAXBAND solutions as starting solutions and perform the TRANSYT optimization.

TABLE 2 Solutions Based on Throughput

Case Name	Model Name	Throughput (vph) Total	Main Street	Side Street	Selected Cycle Length
SKILLMAN AVENUE	TRANSYT	18820	9524	9296	95 Seconds
	BC-TRANSYT	18824	9696	9128	
	MAXBAND	18804	9688	9116	
	COMBAND	18944	9760	9184	
	PASSER II-90	18252	8630	9392	
12TH STREET (AM PEAK)	TRANSYT	17996	6140	11892	90 Seconds
	BC-TRANSYT	17500	6024	11476	
	MAXBAND	16440	5792	10648	
	COMBAND	15560	4812	10748	
	PASSER II-90	15596	5860	11511	
12TH STREET (PM PEAK)	TRANSYT	21020	5740	15280	116 Seconds
	BC-TRANSYT	21224	5764	15460	
	MAXBAND	21020	5940	14074	
	COMBAND	20539	5897	14642	
	PASSER II-90	20988	5756	15232	

delay being reduced as progression bandwidth decreases. This indicates that a trade-off between delay and bandwidth is made usually under well-timed traffic signal systems.

2. The delay on side streets will decrease significantly if some reallocation of green time from arterial to side streets is made, based on the trade-off between the loss of through green bands and the gain of delay saving, especially when the traffic demand on the side streets increases.

3. The optimization feature in TRANSYT usually produces sound system performance in terms of delay even though the bandwidth layout is wiggly or solutions do not have obvious bandwidth. However, the system performance of TRANSYT-

TABLE 3 Disutility Functions

Case Name	Model Name	Total Delay (veh-hr /hr)	Travel Time (min/ mile)	Average Delay (min/ veh)	No. of Stops per trip	Average Speed (mph)	Fuel consum-ption (mpg)	Emiss-ions (kg/ mile-hr)
SKILLMAN AVENUE	TRANSYT	362.38	4.36	3.33	1.42	13.78	10.74	3.613
	BC-TRANSYT	360.46	4.29	3.29	1.30	13.98	10.90	3.549
	MAXBAND	357.71	3.27	4.25	1.34	14.12	10.94	3.561
	COMBAND	363.86	4.31	3.32	1.22	13.92	10.98	3.507
	PASSER II-90	358.54	3.29	4.43	1.28	13.62	10.70	3.464
12TH STREET (AM PEAK)	TRANSYT	298.53	5.31	2.70	1.28	11.30	9.72	1.983
	BC-TRANSYT	313.44	5.88	2.92	1.20	10.28	9.34	1.923
	MAXBAND	367.04	7.05	3.56	1.20	8.54	8.68	1.953
	COMBAND	402.46	8.60	3.90	1.20	6.98	8.22	1.932
	PASSER II-90	355.39	7.05	3.60	1.26	8.54	8.72	1.884
12TH STREET (PM PEAK)	TRANSYT	289.34	4.98	2.15	1.10	12.06	9.38	2.451
	BC-TRANSYT	293.31	5.01	2.16	1.10	11.98	9.36	2.472
	MAXBAND	328.16	5.68	2.44	1.10	10.58	8.94	2.524
	COMBAND	322.33	5.60	2.44	1.10	10.70	9.00	2.523
	PASSER II-90	279.67	4.73	2.06	1.00	12.72	9.72	2.390

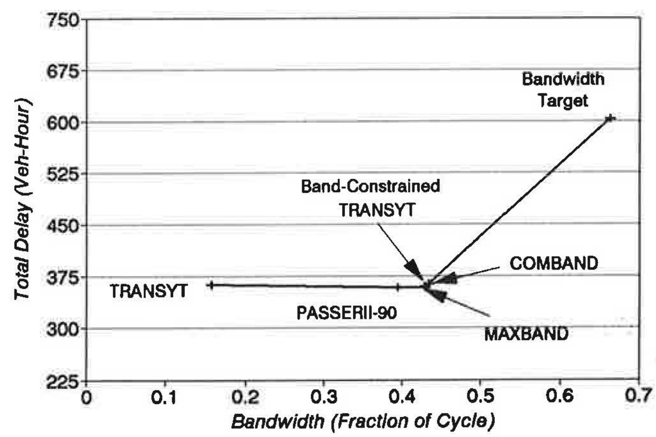

FIGURE 6 Total delay versus bandwidth (Skillman Avenue).

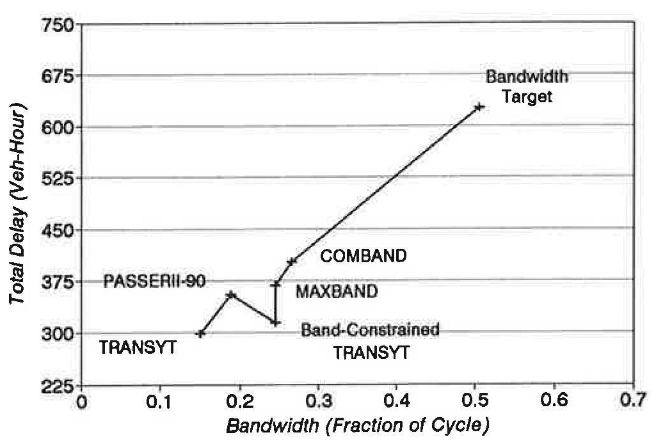

FIGURE 7 Total delay versus bandwidth (12th Street, a.m. peak).

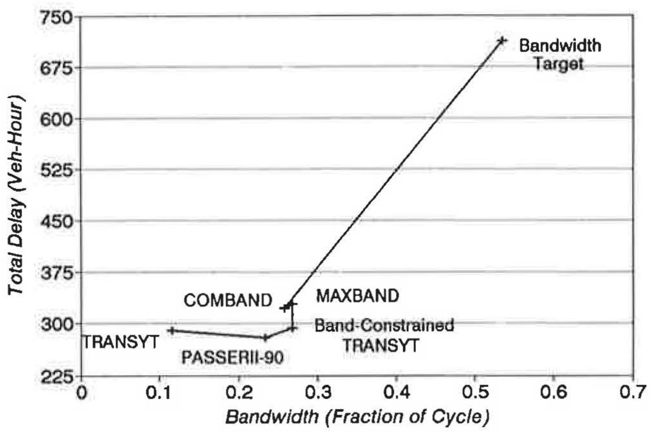

FIGURE 8 Total delay versus bandwidth (12th Street, p.m. peak).

generated solutions is worse than the other models in terms of average number of stops because good progression bandwidth is not provided.

4. Bandwidth-constrained TRANSYT procedures can reduce further delay by subsequently optimizing the MAXBAND initial solutions with the preserved progression bandwidth.

5. PASSER-II 90 shows a tendency to provide a slightly lower bandwidth since it yields more green time for side streets. In general, it produces less delay than the other models.

It is noted that it is not objective to compare these timing solutions by only one criterion. Readers may gain a whole picture of this issue by examining these self-explanatory figures. (Figures 6–8).

CONCLUSIONS

Under a decision-making process, evaluating timing solutions depends on how the decision maker trades off one criterion against the other. In the previous research, two approaches combining delay-minimization and bandwidth-maximization considerations were used to solve the signal coordination problem. One approach is to adjust or fine-tune bandwidth-based timing solutions to further minimize delay by applying a delay-based optimization program. The other approach maximizes bandwidth by modifying the delay-based solutions. These two approaches optimize both timing criteria by either "marrying" two types of programs or adding subprocedures internally or externally. Following concepts similar to those originally proposed in the MITROP model, this paper proposes an alternative approach that provides a viable methodology for simultaneously optimizing two operational criteria delay and progression that are normally conflicting when working with well-timed traffic signal systems.

It is shown that the compromised approach of combining delay/stops and progression bands simultaneously in developing arterial signal timing plans during unsaturated conditions exhibits several advantages over some existing approaches. This approach optimizes the arterial signal timing by performing the trade-off analysis between delay/stop and progression bands criteria simultaneously through MILP method without separately performing the "subprocedures" (such as adjusting the offsets or green splits), although some preprocessing is required. Compared with the basic MAXBAND model, this approach may produce better signal timing plans in terms of delay. The quality of progression bandwidth is still maintained even though some degree of bandwidth may be lost. This approach explicitly optimizes cycle length, system offsets, green splits, and phasing sequences at the same time to achieve a global optimal timing solution. From this methodology, we can either maximize the progression bandwidth at a given user-defined level of service for the side streets or minimize the delay value at some degree of loss in bandwidth.

FUTURE WORK

1. The assumptions and simplifications applied in the underlying model—such as time-stationary flow rate, unsaturated conditions, and no platoon dispersion—can be relaxed further to accommodate real-world situations.

2. The capability of the proposed model can be extended to handle network cases.

3. Multiobjective programming techniques can be applied to deal with signal timing problems having multiple operational criteria.

4. LP-type models suffer from a tremendous computational burden in dealings with a large-scale network problem. Decomposition techniques or other heuristic methods can be introduced to alleviate such a suffering.

ACKNOWLEDGMENTS

This paper is based on research funded by the Texas Department of Transportation (TxDOT) under its Highway Planning and Research Program, whose support is gratefully acknowledged. The authors are solely responsible for all findings, conclusions, and recommendations presented herein.

REFERENCES

1. E. C-P. Chang and C. J. Messer. Minimum Delay Optimization of a Maximum Bandwidth Solution to Arterial Signal Timing. In *Transportation Research Record 1005*, TRB, National Research Council, Washington, D.C., 1985, pp. 80–89.
2. S. L. Cohen and C. C. Liu. The Bandwidth-Constrained TRANSYT Signal-Optimization Program. In *Transportation Research Record 1057*, TRB, National Research Council, Washington, D.C., 1986, pp. 1–9.
3. J. D. C. Little, M. D. Kelson, and N. H. Gartner. MAXBAND: A Program for Setting Signals on Arteries and Triangular Networks. In *Transportation Research Record 795*, TRB, National Research Council, Washington, D.C., 1981, pp. 40–46.
4. N. H. Gartner, J. D. C. Little, and H. Gabbay. Simultaneous Optimization of Offsets, Splits, and Cycle Time. In *Transportation Research Record 596*, 1976, pp. 6–15.
5. N. H. Gartner, J. D. C. Little, and H. Gabbay. Optimization of Traffic Signal Settings by Mixed-Integer Linear Programming—Parts I and II. *Transportation Science*, Vol. 9, No. 4, Nov. 1975, pp. 321–363.

6. E. C-P. Chang, C. J. Messer, and S. L. Cohen. Directional Weighing for Maximal Bandwidth Arterial Signal Optimization Programs. In *Transportation Research Record 1057*, TRB, National Research Council, Washington, D.C., 1986, pp. 10–19.
7. N. H. Gartner, S. F. Assmann, F. Lasaga, and D. L. Hou. A Multi-Band Approach to Arterial Traffic Signal Optimization. *Transportation Research*, Vol. 25B, No. 1, 1991, pp. 55–74.
8. S. Yagar. Minimizing Delay at a Signalized Intersection for Time-Invariant Demand Rates. *Transportation Research*, Vol. 9, 1975, pp. 129–141.
9. *Special Report 209: Highway Capacity Manual*, TRB, National Research Council, Washington, D.C., 1985, Chapter 9.
10. N. A. Chaudhary and T. H. Wu. *Simultaneous Optimization of All Signal Timing Variables*. Working paper. Texas Transportation Institute, Texas A&M University System, College Station, 1991.
11. F. V. Webster. *Traffic Signal Settings*. Road Research Technical Paper 39, 1958.
12. L. Schrage. *User's Manual for Linear, Integer and Quadratic Programming with LINDO*, 3rd ed. The Scientific Press, Redwood City, Calif., 1986.

APPENDIX

NOTATION

$q_{1i}(\bar{q}_{1i}), q_{2i}(\bar{q}_{2i})$ = outbound (inbound) arterial through and sidestreet turning traffic flows of link i (vps);

$b_i(\bar{b}_i)$ = outbound (inbound) bandwidth of link i (cycle);

$r1_{mi}(\overline{r1}_{mi}), r1_{ci}(\overline{r1}_{ci})$ = first component outbound (inbound) main-street red time and sidestreet red time of intersection i (cycle);

$r1_{mi}(\overline{r1}_{mi}), r1_{ci}(\overline{r1}_{ci})$ = second component outbound (inbound) main-street red time and sidestreet red time of intersection i (cycle);

$D1_i(\overline{D1}_i), D2_i(\overline{D2}_i)$ = disutility coefficients associated with each red time variable;

$l_{mi}(\bar{l}_{mi}), l_{ci}(\bar{l}_{ci})$ = outbound (inbound) main-street left-turn green and sidestreet left-turn green time of intersection i;

$w_i(\bar{w}_i)$ = outbound (inbound) leading slack green time outside the bandwidth (cycle);

$\tau_i(\bar{\tau}_i)$ = queue clearance time (cycle);

$t_i(\bar{t}_i)$ = journey time from intersection i to $i + 1$ (i to $1 + i$) (cycle);

$\phi_i(\bar{\phi}_i)$ = internode offsets = time from the beginning of green time at intersection i to beginning of green time at intersection $i + 1$ ($i + 1$ to i) (cycle);

z, C_1, C_2 = reciprocal of common cycle length, lower and upper limits on cycle time;

$e_i, f_i(\bar{e}_i, \bar{f}_i)$ = lower and upper limits on link travel speed from intersection i to $i + 1$ ($i + 1$ to i) (ft/sec);

$g_i, h_i(\bar{g}_i, \bar{h}_i)$ = lower and upper limits on change in link travel speed from link i to $i + 1$ ($i + 1$ to i) (ft/sec); and

n = number of intersections.

FORMULATION

Given

$$q_{1i}(\bar{q}_{1i}),\ q_{2i}(\bar{q}_{2i}),\ D1_i(\overline{D1}_i),\ D2_i(\overline{D2}_i),\ e_i(\bar{e}_i),$$
$$f_i(\bar{f}_i),\ g_i(\bar{g}_i),\ h_i(\bar{h}_i),\ C_1,\ C_2$$

Find

$$b_i,\ \bar{b}_i,\ r1_{mi},\ (\overline{r1}_{mi}),\ r2_{mi},\ (\overline{r2}_{mi}),\ r1_{ci},\ (\overline{r1}_{ci}),$$
$$r2_{ci},\ (\overline{r2}_{ci}),\ l_i,\ \bar{l}_i,\ w_i,\ \bar{w}_i,\ \tau_i,\ \bar{\tau}_i,\ z,\ \delta_i,\ \bar{\delta}_i,\ \lambda_i,\ \bar{\lambda}_i,\ m_i$$

to maximize

$$\sum_i[(q_{1i} - q_{2i})b_i + (\bar{q}_{1i} - \bar{q}_{2i})\bar{b}_i - q_{2i}(r1_{mi} + r2_{mi})$$
$$- \bar{q}_{2i}(\overline{r1}_{mi} + \overline{r2}_{mi}) - (D1_i r1_{mi} + \overline{D1}_i\overline{r1}_{mi} + D2_i r2_{mi}$$
$$+ \overline{D2}_i\overline{r2}_{mi})]$$

subject to

$$w_i + b_i \le 1 - r1_{mi} - r2_{mi} \qquad i = 1, \ldots, n - 1$$

$$w_h + b_i \le 1 - r1_{mh} - r2_{mh} \qquad h = 2, \ldots, n$$

$$\bar{w}_i + \bar{b}_i \le 1 - \overline{r1}_{mi} - \overline{r2}_{mi} \qquad i = 1, \ldots, n - 1$$

$$\bar{w}_h + \bar{b}_i \le 1 - \overline{r1}_{mh} - \overline{r2}_{mh} \qquad h = 2, \ldots, n$$

$$(w_h - \bar{w}_h) + (\bar{w}_i - w_i) + t_i + \bar{t}_i + (r1_{mh} + r2_{mh} - \overline{r1}_{mh}$$
$$- \overline{r2}_{mh}) + (\overline{r1}_{mi} + \overline{r2}_{mi} - r1_{mi} - r2_{mi}) + \delta_h l_h - \delta_i l_i$$
$$- \bar{\delta}_h\bar{l}_h + \bar{\delta}_i\bar{l}_i - (\tau_h + \tau_i) = m_i \qquad i = 1, \ldots, n - 1,$$
$$h = 2, \ldots, n$$

(The loop equations have been modified for releasing green split as variables)

$$zr_i = \min\{1, z/z_{1i}\} \qquad i = 1, \ldots, n$$

$$r_{\min,i} \le r1_{mi}(\overline{r1}_{mi}),\ r1_{ci}(\overline{r1}_{ci}) \le r_{0.5,i} \qquad i = 1, \ldots, n$$

$$0 \le r2_{mi}(\overline{r2}_{mi}), r2_{ci}(\overline{r2}_{ci}) \le (r_{\max,i} - r_{0.5,i}) \qquad i = 1, \ldots, n$$

$$l_i \ge l_{\min,i} \qquad i = 1, \ldots, n$$

$\tau_i = \max[que_i, 0]$ (queue clearance time constraints)

$$i = 1, \ldots, n - 1$$

$$r1_{mi} + r2_{mi} + (\text{Webster split on main}) \cdot zr_i - \bar{l}_{mi}$$
$$\le 1 \qquad i = 1, \ldots, n$$

$$\overline{r1}_{mi} + \overline{r2}_{mi} + (\text{Webster split on main}) \cdot zr_i - l_{mi}$$
$$\le 1 \qquad i = 1, \ldots, n$$

$$r1_{ci} + r2_{ci} + (\text{Webster split on side}) \cdot zr_i - \bar{l}_{ci}$$

$$\leq 1 \qquad i = 1, \ldots, n$$

$$\overline{r1}_{ci} + \overline{r2}_{ci} + (\text{Webster split on side}) \cdot zr_i - l_{ci}$$

$$\leq 1 \qquad i = 1, \ldots, n$$

$$l_{mi} - \overline{r1}_{mi} - \overline{r2}_{mi} = \bar{l}_{mi} - r1_{mi} - r2_{mi} \qquad i = 1, \ldots, n$$

$$l_{ci} - \overline{r1}_{ci} - \overline{r2}_{ci} = \bar{l}_{ci} - r1_{ci} - r2_{ci} \qquad i = 1, \ldots, n$$

$$-l_{mi} + \overline{r1}_{mi} + \overline{r2}_{mi} - \bar{l}_{ci} + r1_{ci} + r2_{ci} = 1$$

$$i = 1, \ldots, n$$

$$(d_i/f_i)z \leq t_i \leq (d_i/e_i)z \quad i = 1, \ldots, n - 1$$

$$(\bar{d}_i/\bar{f}_i)z \leq \bar{t}_i \leq (\bar{d}_i/\bar{e}_i)z \quad i = 1, \ldots, n - 1$$

$$(d_i/h_i)z \leq (d_i/d_{i+1})t_{i+1} - t_i \leq (d_i/g_i)z \quad i = 1, \ldots, n - 2$$

$$(\bar{d}_i/\bar{h}_i)z \leq (\bar{d}_i/\bar{d}_{i+1})\bar{t}_{i+1} - \bar{t}_i \leq (\bar{d}_i/\bar{g}_i)z \quad i = 1, \ldots, n - 2$$

(The lower and upper limits on journey time have been modified in COMBAND considering the degree of saturation of each link.)

$$1/C_2 \leq z \leq 1/C_1$$

m_i = integer variables

δ_i, $\bar{\delta}_i$ = 0 − 1 binary variables

This paper does not represent the official views of policies of TxDOT or FHWA. This document does not constitute a standard, specification, or regulation.

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