PERGAMON

Mathematical and Computer Modelling 35 (2002) 657–671

MATHEMATICAL AND COMPUTER MODELLING

www.elsevier.com/locate/mcm

Arterial-Based Control of Traffic Flow in Urban Grid Networks

N. H. Gartner and C. Stamatiadis
Department of Civil and Environmental Engineering
University of Massachusetts, Lowell
Lowell, MA 01854, U.S.A.
<Nathan_Gartner><Chronis_Stamatiadis>@uml.edu

Abstract—Urban networks are typically composed of a grid of arterial streets. Optimal control of the traffic signals in the grid system is essential for the effective operation of the network. In this paper, we present mathematical programming models for the development of optimal arterial-based progression schemes. Such schemes are widely used for traffic signal control in arterial streets. Under such a scheme, a continuous green band is provided in each direction along the artery at the desired speed of travel to facilitate the movement of through traffic along the arterial. Traditional schemes consist of uniform-width progressions. New approaches generate variable bandwidth progressions in which each directional road section is allocated an individually weighted band that can be adapted to the prevailing traffic flows on that link. Mixed-integer linear programming is used for the optimization. Simulation results indicate that this method can produce considerable gains in performance when compared with traditional progression methods. By introducing efficient computational techniques, this method also lends itself to a natural extension for incorporation in a dynamic traffic management system.

Keywords—Traffic flow, Traffic control, Traffic signals, Bandwidth progression, Mixed-integer linear programming.

1. INTRODUCTION

The predominant flow of traffic in urban street networks is along arterial streets. Optimal control of the traffic signals along these arteries is essential for the effective operation of the grid network. Coordination of traffic lights along arterial streets provides numerous advantages as indicated in the traffic engineering literature [1].

(1) A higher level of traffic service is provided in terms of higher overall speed and reduced number of stops.
(2) Traffic flows more smoothly, often with an improvement in capacity due to reduced headways.
(3) Vehicle speeds are more uniform because there is no incentive to travel at excessively high speeds to reach a signalized intersection within a green interval that is not in step. Also, the slow driver is encouraged to speed up in order to avoid having to stop for a red light.
(4) There are fewer accidents because the platoons of vehicles arrive at each signal when it is green, thereby reducing the possibility of red signal violations and rear-end collisions.

0895-7177/02/$ - see front matter Typeset by $\mathcal{AMS}$-TEX
PII: S0895-7177(01)00188-1

(5) Greater obedience to the signal commands is obtained from both motorists and pedestrians. The motorist tries to keep within the green interval, and the pedestrian stays at the curb because the vehicles are more tightly spaced.
(6) Through traffic tends to stay on the arterial streets instead of diverting onto parallel minor streets in search of alternative routes.

There are two basic approaches to computing arterial signal timings [2]:

- maximize the bandwidth of the progression, or
- minimize overall delays and stops.

Arterial progression schemes are widely used for signal coordination. The conceptual basis for the progression design is that traffic signals tend to group vehicles into a "platoon" with more uniform headways than would otherwise occur. The platooning effect is most prevalent on major streets with signalized intersections at frequent intervals. It is desirable, in these circumstances, to encourage platooning so that continuous movement (or progression) of vehicle platoons through successive traffic lights can be maintained. The signal timings, in this case, are designed to maximize the width of continuous green bands in both directions along the artery at the expected or recommended speed of travel. Although maximizing the bandwidth is only a surrogate performance measure, it has been shown to provide significant benefits in increased travel speeds and smoothness of traffic flow and is the preferred approach in many countries [3,4]. In general, such signal systems operate best when the main-street flow is predominantly through traffic and when the number of vehicles turning onto the main street is small compared with the through volume. The second approach uses direct performance models to minimize delay, stops, or other measures of disutility. Due to the complexity of these models and the high degree of nonlinearity involved, they are generally not amenable to solution by mathematical programming methods, and heuristic approaches have to be used. As a consequence, the resulting solutions are nonoptimal. In this paper, we concentrate on models of the first kind.

The first optimization model used for arterial bandwidth maximization was developed by Morgan and Little using a combinatorial optimization scheme [5]. A more advanced model, using mixed-integer linear programming, was later formulated by Little [6]. Advances in optimization techniques and computational capabilities have steadily increased the sophistication and versatility of the traffic signal optimization models. In this paper, we introduce, first, single arterial bandwidth optimization models using both uniform and variable-width bandwidth progressions. These models are then extended to the optimization of arterial grid networks. Heuristic decomposition approaches are developed to speed-up the computation and to make it suitable for on-line implementation as well as for combination with traffic assignment and routing models.

The models presented below use mixed-integer linear programming (MILP) for the optimization, and are presented with progressively increasing complexity. MILP-1 is the basic uniform-width bandwidth maximization model for a single arterial. MILP-2 introduces a multiband, multiweight approach that generates variable-width two-way progression bands along the arterial. MILP-3 extends the previous models to a grid network of intersecting arterials.

2. THE BASIC BANDWIDTH MAXIMIZATION PROBLEM

The geometric relations for the uniform bandwidth model are shown in Figure 1. Consider an arterial having n signalized intersections. Let S_i denote the i^{th} signal on the arterial and L_i denote the i^{th} link (between signals i and $i+1$) of the arterial, with $i = 1, \dots, n_j$. All time variables are defined in units of the cycle time. The following variables are defined:

C = cycle time (sec);

$b\ (\bar{b})$ = outbound (inbound) bandwidth;

$r_i\ (\bar{r}_i)$ = outbound (inbound) red time at S_i;

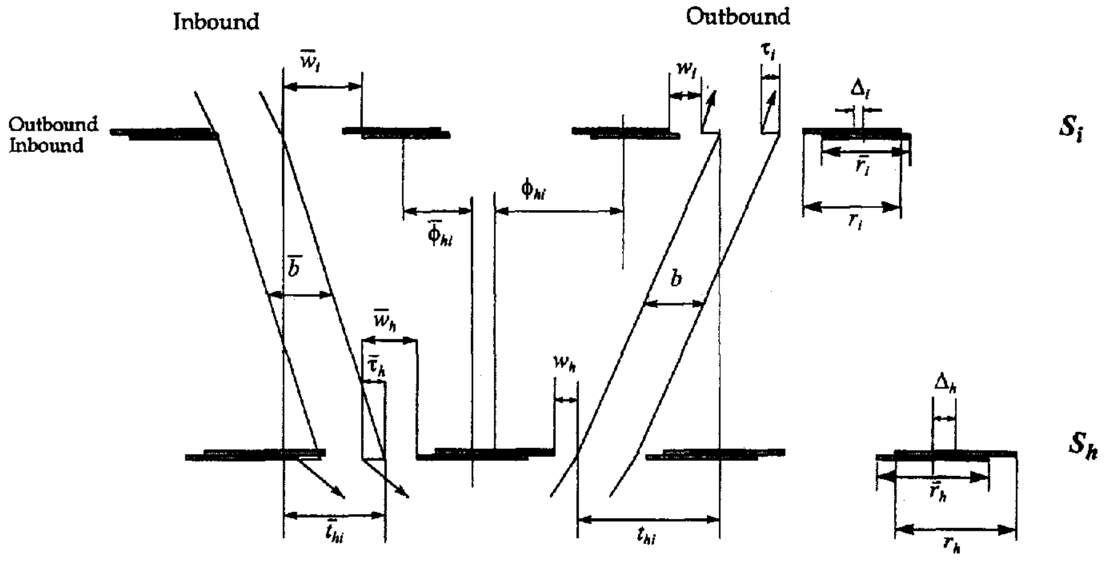

Figure 1. Time space diagram for MILP-1.

w_i ($\bar{w}_i$) = interference variables, time from right (left) side of red at S_i to left (right) side of outbound (inbound) green band;

t_{hi} ($\bar{t}_{hi}$) = travel time from S_i to S_h in the outbound (inbound) direction;

d_{hi} ($\bar{d}_{hi}$) = distance from S_i to S_h in the outbound (inbound) direction;

ϕ_{hi} ($\bar{\phi}_{hi}$) = internode offsets, time from the center of the outbound (inbound) red at S_h to the center of the outbound (inbound) red at S_i;

Δ_i = directional node phase shift, time from center of $\bar{r}_i$ to nearest center of r_i;

τ_i ($\bar{\tau}_i$) = queue clearance time for advancement of outbound (inbound) bandwidth at S_i to clear turning-in traffic before arrival of main-street platoon;

V_i ($\bar{V}_i$) = outbound (inbound) progression speed on link L_i (m/sec).

In the case of uniform bandwidths, the objective function has the following form:

$$\text{Maximize } b + k \cdot \bar{b}, \tag{1}$$

where k is the target ratio of inbound to outbound bandwidth taken as the ratio of total inbound to total outbound volumes along the arterial.

The first set of constraints that we introduce are the *directional interference constraints*. These constraints make sure that the progression bands use only the available green time and do not infringe upon any of the red times. From Figure 1, we see that

$$w_i + b \leq 1 - r_i, \tag{2a}$$

$$\bar{w}_i + \bar{b} \leq 1 - \bar{r}_i. \tag{2b}$$

Next, we calculate the *arterial-loop integer constraint*. This constraint is due to the fact that the signals of the arterial are synchronized; i.e., they operate in a common cycle time. If we start at the center of the outbound red at S_i and proceed along a loop consisting of the following points:

center of outbound red at $S_i \rightarrow$
center of inbound red at $S_i \rightarrow$
center of inbound red at $S_h \rightarrow$
center of outbound red at S_h,

we must end up at a point that is removed an integral number of cycle times from the point of departure. Summing algebraically the appropriate internode offsets and the directional node phase shift along the loop, we obtain

$$\phi_{hi} + \bar{\phi}_{hi} + \Delta_h - \Delta_i = v_{hi}, \tag{3}$$

where v_{hi} is the corresponding loop integer variable.

From Figure 1, we observe the following identities:

$$\phi_{hi} + \frac{r_i}{2} + w_i + \tau_i = \frac{r_h}{2} + w_h + t_{hi} \tag{4a}$$

and

$$\bar{\phi}_{hi} + \frac{\bar{r}_i}{2} + \bar{w}_i + \bar{\tau}_i = \frac{\bar{r}_h}{2} + \bar{w}_h + \bar{t}_{hi}. \tag{4b}$$

Substituting (4) into (3) to eliminate ϕ and $\bar{\phi}$ gives

$$t_{hi} + \bar{t} + hi + \frac{r_h + \bar{r}_h}{2} + (w_h + \bar{w}_h) - \frac{r_i + \bar{r}_i}{2} - (w_i + \bar{w}_i) - (\tau_i + \bar{\tau}_h) + \Delta_h - \Delta_i = v_{hi}. \tag{5}$$

So far we have assumed that S_i follows S_h in the outbound direction, but this restriction is not necessary. Letting x represent any of the variables t, $\bar{t}$, v, ϕ, $\bar{\phi}$, then the following relationship is satisfied:

$$x_{hj} = x_{hi} + x_{ij}, \tag{6}$$

whence, by setting $h = j$, we obtain $x_{hi} = -x_{ih}$. Thus, equations (3) and (4) hold for arbitrary S_h and S_i.

To simplify notation, we number the signals sequentially from 1 to n in the outbound direction, and we define $x_i = x_{i,i+1}$. We can now rewrite (5) as follows:

$$t_i + \bar{t}_i + \frac{r_i + \bar{r}_i}{2} + (w_i + \bar{w}_i) - \frac{r_{i+1} + \bar{r}_{i+1}}{2} - (w_{i+1} + \bar{w}_{i+1}) - (\tau_{i+1} + \bar{\tau}_i) + \Delta_i - \Delta_{i+1} = v_i. \tag{7}$$

The common signal cycle time C (seconds) and the link specific progression speed V_i ($\bar{V}_i$) (m/sec) are optimizable variables as well. This introduces considerable flexibility in the calculation of the best arterial progression. Each of these variables is constrained by upper and lower limits. In addition, changes in speed from one link to the next can also be limited. Let the limits be as follows:

$$\begin{aligned} C_1, C_2 &= \text{lower and upper bounds on cycle length;} \\ (e_i, f_i), (\bar{e}_i, \bar{f}_i) &= \text{lower and upper bounds on outbound (inbound) speed } V_i\ (\bar{V}_i)\ \text{(m/sec);} \\ (g_i, h_i), (\bar{g}_i, \bar{h}_i) &= \text{lower and upper bounds on change in outbound (inbound)} \\ &\ \text{speed } V_i\ (\bar{V}_i)\ \text{(m/sec).} \end{aligned}$$

To obtain constraints that are linear in the decision variables, we define the inverse of the cycle time, i.e., the signal frequency $z = 1/C$ (cycles/second), such that

$$\frac{1}{C_2} \le z \le \frac{1}{C_1}. \tag{8}$$

The progression speeds and the changes in speeds are also expressed in terms of their reciprocals

$$\frac{1}{f_i} \le \frac{1}{V_i} \le \frac{1}{e_i} \quad \text{and} \quad \frac{1}{\bar{f}_i} \le \frac{1}{\bar{V}_i} \le \frac{1}{\bar{e}_i}, \tag{9}$$

$$\frac{1}{h_i} \le \frac{1}{V_{i+1}} - \frac{1}{V_i} \le \frac{1}{g_i} \quad \text{and} \quad \frac{1}{\bar{h}_i} \le \frac{1}{\bar{V}_{i+1}} - \frac{1}{\bar{V}_i} \le \frac{1}{\bar{g}_i}. \tag{10}$$

Using the notation $d_i = d_{i,i+1}$, we obtain

$$t_i = \frac{d_i}{V_i} z \qquad \text{and} \qquad \bar{t}_i = \frac{\bar{d}_i}{\bar{V}_i} z.$$

Substituting these expressions above and manipulating the variables, we obtain

$$\frac{d_i}{f_i} z \leq t_i \leq \frac{d_i}{e_i} z \qquad \text{and} \qquad \frac{\bar{d}_i}{\bar{f}_i} z \leq \bar{t}_i \leq \frac{\bar{d}_i}{\bar{e}_i} z, \tag{11}$$

$$\frac{d_i}{h_i} z \leq \frac{d_i}{d_{i+1}} t_{i+1} - t_i \leq \frac{d_i}{g_i} z \qquad \text{and} \qquad \frac{\bar{d}_i}{\bar{h}_i} z \leq \frac{\bar{d}_i}{\bar{d}_{i+1}} \bar{t}_{i+1} - \bar{t}_i \leq \frac{\bar{d}_i}{\bar{g}_i} z. \tag{12}$$

Thus, t_i, $\bar{t}_i$, and z are decision variables which, once known, determine the progression speeds. The complete MILP-1 is given below.

MILP-1. Find b, $\bar{b}$, z, w_i, $\bar{w}_i$, t_i, $\bar{t}_i$, v_i **to**

Maximize $b + k \cdot \bar{b}$ **subject to**

$$\frac{1}{C_2} \leq z \leq \frac{1}{C_1},$$

$$w_i + b \leq 1 - r_i \quad \text{and} \quad \bar{w}_i + \bar{b} \leq 1 - \bar{r}_i, \qquad i = 1, \dots, n,$$

$$t_i + \bar{t}_i + \frac{r_i + \bar{r}_i}{2} + (w_i + \bar{w}_i) - \frac{r_{i+1} + \bar{r}_{i+1}}{2} - (w_{i+1} + \bar{w}_{i+1}) - (\tau_{i+1} + \bar{\tau}_i) + \Delta_i - \Delta_{i+1} = v_i, \qquad i = 1, \dots, n-1,$$

$$\frac{d_i}{f_i} z \leq t_i \leq \frac{d_i}{e_i} z \quad \text{and} \quad \frac{\bar{d}_i}{\bar{f}_i} z \leq \bar{t}_i \leq \frac{\bar{d}_i}{\bar{e}_i} z, \qquad i = 1, \dots, n-1,$$

$$\frac{d_i}{h_i} z \leq \frac{d_i}{d_{i+1}} t_{i+1} - t_i \leq \frac{d_i}{g_i} z \quad \text{and} \quad \frac{\bar{d}_i}{\bar{h}_i} z \leq \frac{\bar{d}_i}{\bar{d}_{i+1}} \bar{t}_{i+1} - \bar{t}_i \leq \frac{\bar{d}_i}{\bar{g}_i} z, \qquad i = 1, \dots, n-1,$$

$$b, \bar{b}, z, w_i, \bar{w}_i, t_i, \bar{t}_i \geq 0 \quad \text{and} \quad v_i \text{ integer}.$$

In addition to bandwidths and interference variables, this program also calculates optimal cycle length $(1/z)$ and progression speeds. To calculate the offsets, green splits need to be given or, alternatively, the user can provide traffic volume and capacity information for each intersection and the program will calculate the splits. MILP-1 is essentially the current version of the MAXBAND program provided developed by Little *et al.* [7]. It involves $(11n - 10)$ constraints and $(4n + 1)$ continuous variables and $n - 1$ unrestricted integer variables, not counting slack variables. In addition, if an option for phase sequencing is activated (not included here), an additional $2n$ integer variables and constraints may be required. An example of a uniform bandwidth solution obtained by MILP-1 problem is given in Figure 2. The scheme contains signals with multiphase sequences.

3. THE VARIABLE BANDWIDTH PROBLEM

In this section, we describe a more versatile multiband/multiweight approach for the bandwidth maximization problem. In this case, a different bandwidth is defined for each directional road section of the arterial that is individually weighted with respect to its contribution to the overall objective function. While the band is continuous, its width can vary and adapt to the prevailing traffic flow on each link. In this way, the formulation is sensitive to varying traffic conditions and can tailor the progression scheme to the different possible traffic flow patterns. The user can still choose a uniform bandwidth progression if desired, but this is only one of the many possible options. Referring to the geometry in Figure 3, we define the following variables:

$b_i(\bar{b}_i) =$ outbound (inbound) bandwidth of link i; there is now a specific band for each link L_i;

$w_i(\bar{w}_i) =$ the time from right (left) side of red at S_i to the centerline of the outbound (inbound) green band; the reference point at each signal has been moved from the edges to the centerline of the band.

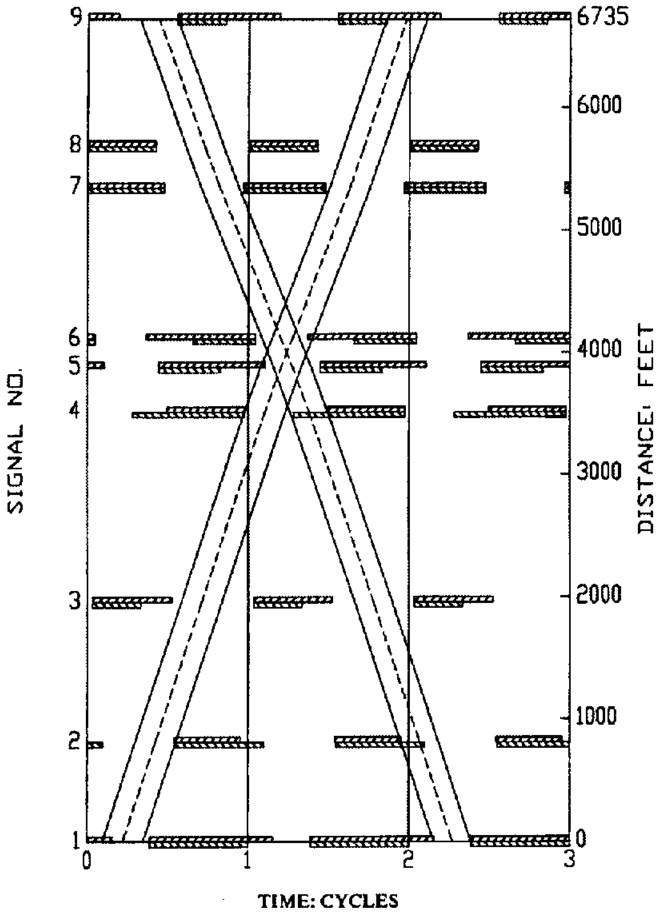

Figure 2. Time space diagram for a uniform bandwidth progression scheme generated by MILP-1.

The following constraints apply in the outbound directions at signal S_i:

$$w_i + \frac{b_i}{2} \leq 1 - r_i \quad \text{and} \quad w_i + \frac{b_i}{2} \geq 0. \tag{13}$$

The pair of constraints can be combined as follows:

$$\frac{b_i}{2} \leq w_i \leq (1 - r_i) - \frac{b_i}{2}. \tag{14a}$$

A similar relation must be observed at S_{i+1}, since band b_i must be constrained at both ends,

$$\frac{b_i}{2} \leq w_{i+1} \leq (1 - r_{i+1}) - \frac{b_i}{2}. \tag{14b}$$

Corresponding relationships exist in the inbound direction (marked by a bar on all variables),

$$\frac{\bar{b}_i}{2} \leq \bar{w}_i \leq (1 - r_i) - \frac{\bar{b}_i}{2}, \tag{14c}$$

$$\frac{\bar{b}_i}{2} \leq \bar{w}_{i+1} \leq (1 - \bar{r}_{i+1}) - \frac{\bar{b}_i}{2}. \tag{14d}$$

These constraints are referred to as the *directional interference constraints*. The time reference points are redefined to the centerline of the bands (or, the *progression* line), rather than the

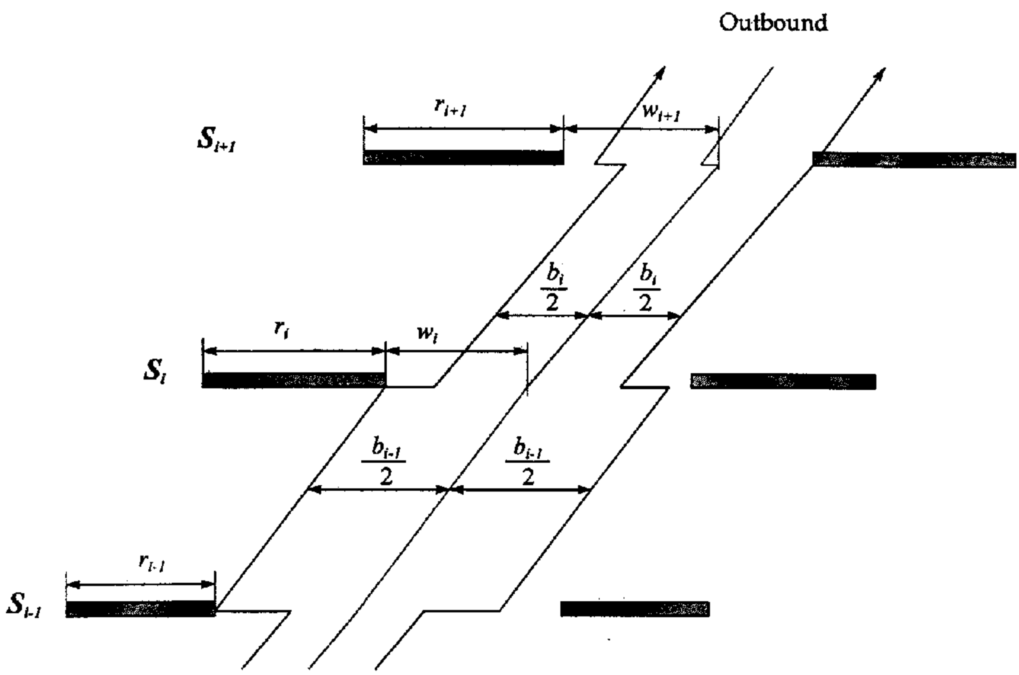

Figure 3. Geometric relations for MILP-2.

edges. The *loop integer constraint* given by equation (7) remains unchanged and so do the travel time and speed-change constraints.

The most important change occurs in the objective function. Since the bands are link-specific, they can be weighted disaggregately to reflect user-defined traffic performance objectives which can be individually weighted for each link of the arterial. The objective function now has the following form:

$$\text{Maximize } \frac{1}{n-1} \sum_{i=1}^{n} a_i \cdot b_i + \bar{a}_i \cdot \bar{b}_i, \tag{15}$$

where a_i and $\bar{a}_i$ are the link specific weighting coefficients for the outbound and inbound directions, respectively. There is a multitude of possible options for choosing the weighting coefficients in (15). Some of the more common expressions are as follows:

$$a_i \left(\frac{q_i}{s_i}\right)^p \quad \text{and} \quad \bar{a}_i = \left(\frac{\bar{q}_i}{\bar{s}_i}\right)^p, \tag{16}$$

where

$q_i(\bar{q}_i) =$ outbound (inbound) directional flow rate on link L_i; either the total or the through volume can be used;

$s_i(\bar{s}_i) =$ saturation flow rate outbound (inbound) directional volume on link V_{ij}; either the total flow rate or the through flow rate can be used;

$p =$ an integer exponent; the values 0, 1, 2, and 4 were used.

To obtain an objective function value that is consistent with those used previously, we normalize the weighting coefficients to obtain

$$\sum_{i=1}^{n-1} a_i = n-1 \quad \text{and} \quad \sum_{i=1}^{n-1} \bar{a}_i = n-1.$$

The multiband/multiweight optimization program is summarized in MILP-2.

Table 1. Constraints and variables for MILP-1 compared to MILP-2.

Constraint	Variable	MIPLP-1		MILP-2	
		No. of Constraints	No. of Variables	No. of Constraints	No. of Variables
Bands	$b_1, \bar{b}_i$	—	2	—	$2(n-1)$
Cyclic rate	z	2	1	2	1
Interference	$w_i, \bar{w}_i$	$2n$	$2n$	$8(n-1)$	$2n$
Loop integer	v_1	$n-1$	$n-1$	$n-1$	$n-1$
Travel times (speed)	$t_i, \bar{t}_i$	$4(n-1)$	$2(n-1)$	$4(n-1)$	$2(n-1)$
Speed changes	—	$4(n-2)$	—	$4(n-2)$	—
Totals		$11n-10$	$5n$	$18n-20$	$7n-4$
Example: $n=5$		45	25	70(+55%)	31(+24%)
$n=10$		100	50	160(+60%)	66(+32%)

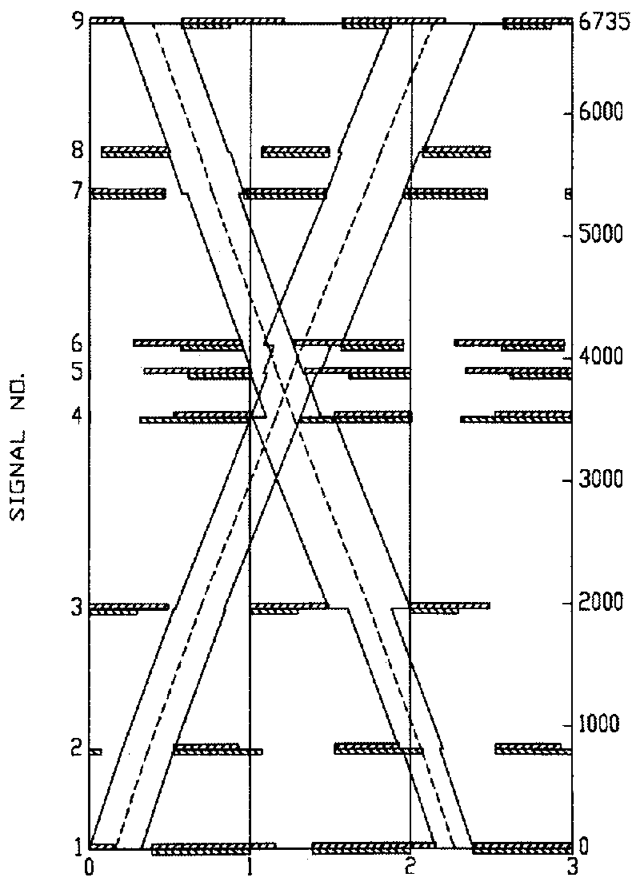

Figure 4. Time space diagram for a variable bandwidth progression scheme obtained by MILP-2.

MILP-2. Find b_i, $\bar{b}_i$, z, w_i, $\bar{w}_i$, t_i, $\bar{t}_i$, v_i to

$$\textbf{Maximize } \frac{1}{n-1}\sum_{i=1}^{n} a_i \cdot b_i + \bar{a}_i \cdot \bar{b}_i \textbf{ subject to}$$

$$\frac{1}{C_2} \le z \le \frac{1}{C_1},$$

$$\frac{b_i}{2} \le w_i \le (1 - r_i) - \frac{b_i}{2}, \qquad \frac{b_i}{2} \le w_{i+1} \le (1 - r_{i+1}) - \frac{b_i}{2},$$
$$\frac{\bar{b}_i}{2} \le \bar{w}_i \le (1 - \bar{r}_i) - \frac{\bar{b}_i}{2} \quad \text{and} \quad \frac{\bar{b}_i}{2} \le \bar{w}_{i+1} \le (1 - \bar{r}_{i+1}) - \frac{\bar{b}_i}{2}, \qquad i = 1, \dots, n-1,$$
$$t_i + \bar{t}_i + \frac{r_i + \bar{r}_i}{2} + (w_i + \bar{w}_i) - \frac{r_{i+1} + \bar{r}_{i+1}}{2} - (w_{i+1} + \bar{w}_{i+1}) - (\tau_{i+1} + \bar{\tau}_i) + \Delta_i - \Delta_{i+1} = v_i, \qquad i = 1, \dots, n-1,$$
$$\frac{d_i}{f_i} z \le t_i \le \frac{d_i}{e_i} z \quad \text{and} \quad \frac{\bar{d}_i}{\bar{f}_i} z \le \bar{t}_i \le \frac{\bar{d}_i}{\bar{e}_i} z, \qquad i = 1, \dots, n-1,$$
$$\frac{d_i}{h_i} z \le \frac{d_i}{d_{i+1}} t_{i+1} - t_i \le \frac{d_i}{g_i} z \quad \text{and} \quad \frac{\bar{d}_i}{\bar{h}_i} z \le \frac{\bar{d}_i}{\bar{d}_{i+1}} \bar{t}_{i+1} - \bar{t}_i \le \frac{\bar{d}_i}{\bar{g}_i} z, \qquad i = 1, \dots, n-1,$$
$$b, \bar{b}_i, z, w_i, \bar{w}_i, t_i, \bar{t}_i \ge 0, \quad \text{and} \quad v_i \text{ integer}.$$

MILP-2 involves $(18n - 20)$ constraints and $(6n - 3)$ continuous variables, and $n - 1$ unrestricted integer variables, not counting slack variables. Green or red splits are determined the same as in MILP-1. Table 1 gives a comparison of the number of constraints and variables for MILP-1 and MILP-2. The increased decision capabilities of MILP-2 require a corresponding increase in the size of the mathematical program. The formulation given in MILP-2 corresponds to the MULTIBAND optimization program developed by Gartner *et al.* [8]. The variable bandwidth solution for the same arterial as for MILP-1 (Figure 1) is shown in Figure 4.

4. THE NETWORK PROBLEM

The MILP-2 problem can be extended for controlling traffic flow in a network of intersecting arterials. This represents a significantly more challenging problem: progressions must be provided on all the arterials of the network, simultaneously. The principal difficulty is due to the requirement that the progression bands of intersecting arterials cannot overlap in time at any junction of the network.

Consider a network of m arterials with each arterial having n_j signalized intersections. Let S_{ij} denote the i^{th} signal on the j^{th} arterial of the network and L_{ij} denote the i^{th} link (between signals i and $i + 1$) of the j^{th} arterial, with $j = 1, \dots, m$ and $i = 1, \dots, n_j$. The variables presented in MILP-1 and MILP-2 are redefined as follows:

b_{ij} $(\bar{b}_{ij})$ = outbound (inbound) bandwidth of link i on arterial j;

w_{ij} $(\bar{w}_{ij})$ = interference variables, the time from right (left) side of red at S_{ij} to the centerline of the outbound (inbound) green band;

r_{ij} $(\bar{r}_{ij})$ = outbound (inbound) red time at S_{ij};

t_{ij} $(\bar{t}_{ij})$ = travel time on link i of arterial j in the outbound (inbound) direction;

d_{ij} $(\bar{d}_{ij})$ = length of link i of arterial j in the outbound (inbound) direction;

$\phi_{(ij),(kj)}$ $(\bar{\phi}_{(ij),(kj)})$ = internode offsets, time from the center of the outbound (inbound) red at S_{ij} to the center of the outbound (inbound) red at S_{kj};

Δ_{ij} = directional node phase shift, time from center of $\bar{r}_{ij}$ to nearest center of r_{ij};

τ_{ij} $(\bar{\tau}_{ij})$ = queue clearance time for advancement of outbound (inbound) bandwidth at S_{ij} to clear turning-in traffic before arrival of main-street platoon;

V_{ij} $(\bar{V}_{ij})$ = outbound (inbound) progression speed on link L_{ij} (m/sec);

$e_{ij}, f_{ij}, (\bar{e}_{ij}, \bar{f}_{ij})$ = lower and upper bounds on outbound (inbound) speed V_{ij} $(\bar{V}_{ij})$ (m/sec);

$g_{ij}, h_{ij}, (\bar{g}_{ij}, \bar{h}_{ij})$ = lower and upper bounds on change in outbound (inbound) speed V_{ij} $(\bar{V}_{ij})$ (m/sec).

In addition to the above variables, the intranode offset is defined as follows:

$\omega_{(ij),(kl)}$ $(\bar{\omega}_{(ij),(kl)})$ = intranode offsets, the time from the center of the inbound (outbound) red at i^{th} node of the j^{th} arterial (S_{ij}), to the center of the inbound (outbound) red in the crossing direction at the same node which is also the k^{th} node of the l^{th} arterial (S_{kl}).

The directional interference constraints, the arterial loop integer constraint, and the travel time and speed-change constraints remain unchanged and are applied, in turn, for each arterial. A new set of constraints, though, is required for the synchronization of all signals. Similar to the arterial loop-integer constraints (equation (3)), for any closed loops of the network consisting of more than two links, the summation of *internode* and *intranode offsets* around the loop of intersecting arterials must be an integer multiple of the cycle time. This results in the formulation of the *network loop-integer constraints* as follows:

$$\begin{aligned} &\phi_{(ia),(i+1,a)} + \omega_{(i+1,a),(jb)} + \phi_{(jb),(j+1,b)} + \omega_{(j+1,b),(kc)} + \phi_{(kc),(k+1,c)} \\ &\quad + \omega_{(k+1,c),(ld)} + \phi_{(ld),(l+1,d)} + \omega_{(ld),(ia)} = \mu_N, \end{aligned} \tag{17}$$

where a, b, c, and d are intersecting arterials as in Figure 5 and μ_N is the integer variable of the N^{th} network loop. The number of network loop constraints and the choice of a fundamental set of loops is given by Gartner [9].

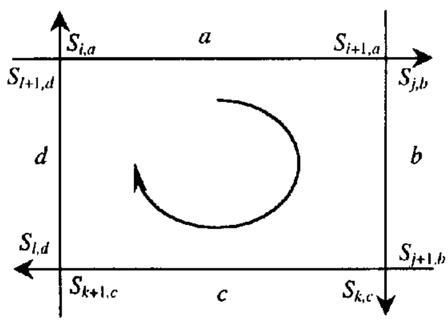

Figure 5. Network loop.

The network optimization problem is summarized in MILP-3.

MILP-3. Find b_{ij}, $\bar{b}_{ij}$, z, w_{ij}, $\bar{w}_{ij}$, t_{ij}, $\bar{t}_{ij}$, v_{ij}, μ_i to

$$\textbf{Maximize} \sum_{j=1}^{m} \frac{1}{n-1} \sum_{i=1}^{n} a_{ij} \cdot b_{ij} + \bar{a}_{ij} \cdot \bar{b}_{ij} \textbf{ subject to}$$

$$\frac{1}{C_2} \le z \le \frac{1}{C_1},$$

$$\frac{b_{ij}}{2} \le w_{ij} \le (1 - r_{ij}) - \frac{b_{ij}}{2}, \qquad \frac{b_{ij}}{2} \le w_{i+1,j} \le (1 - r_{i+1,j}) - \frac{b_{ij}}{2},$$

$$\frac{\bar{b}_{ij}}{2} \le \bar{w}_{ij} \le (1 - \bar{r}_{ij}) - \frac{\bar{b}_{ij}}{2} \quad \text{and} \quad \frac{\bar{b}_{i,j}}{2} \le \bar{w}_{i+1,j} \le (1 - \bar{r}_{i+1,j}) - \frac{\bar{b}_{ij}}{2},$$
$$i = 1, \dots, n_j - 1; \quad j = 1, \dots, m,$$

$$\begin{aligned} &t_{ij} + \bar{t}_{ij} + \frac{r_{ij} + \bar{r}_{ij}}{2} + (w_{ij} + \bar{w}_{ij}) - \frac{r_{i+1,j} + \bar{r}_{i+1,j}}{2} - (w_{i+1,j} + \bar{w}_{i+1,j}) \\ &\quad - (\tau_{i+1,j} + \bar{\tau}_{i,j}) + \Delta_{i,j} - \Delta_{i+1,j} = v_{i,j}, \qquad i = 1, \dots, n_j - 1; \quad j = 1, \dots, m, \end{aligned}$$

$$\begin{aligned} &\phi_{(ia),(i+1,a)} + \omega_{(i+1,a),(jb)} + \phi_{(jb),(j+1,b)} + \omega_{(j+1,b),(kc)} + \phi_{(kc),(k+1,c)} \\ &\quad + \omega_{(k+1,c),(ld)} + \phi_{(ld),(l+1,d)} + \omega_{(ld),(ia)} = \mu_N, \end{aligned}$$

$$i = 1, \ldots, n_a - 1; \quad j = 1, \ldots, n_b - 1; \quad k = 1, \ldots, n_c - 1; \quad l = 1, \ldots, n_d - 1;$$

a, b, c, d arterials forming a fundamental network loop;

$$\frac{d_{ij}}{f_{ij}} z \le t_{ij} \le \frac{d_{ij}}{e_{ij}} z \quad \text{and} \quad \frac{\bar{d}_{ij}}{\bar{f}_{ij}} z \le \bar{t}_{ij} \le \frac{\bar{d}_{ij}}{\bar{e}_{ij}} z, \qquad i = 1, \ldots, n_j - 1; \quad j = 1, \ldots, m,$$

$$\frac{d_{ij}}{h_{ij}} z \le \frac{d_{ij}}{d_{i+1,j}} t_{i+1,j} - t_{ij} \le \frac{d_{ij}}{g_{ij}} z \quad \text{and} \quad \frac{\bar{d}_{ij}}{\bar{h}_{ij}} z \le \frac{\bar{d}_{ij}}{\bar{d}_{i+1,j}} \bar{t}_{i+1,j} - \bar{t}_{ij} \le \frac{\bar{d}_{ij}}{\bar{g}_{ij}} z,$$
$$i = 1, \ldots, n_j - 1; \quad j = 1, \ldots, m,$$

$$b_{ij}, \bar{b}_{ij}, z, w_{ij}, \bar{w}_{ij}, t_{ij}, \bar{t}_{ij} \ge 0, \quad \text{and} \quad v_{ij},\ \mu_i \text{ integer.}$$

For an $m \times n$ closed grid network ($m \times n$ intersections and $m + n$ arterials), MILP-3 may involve up to $(23mn - 14m - 14n + 1)$ constraints and $(12mn - 6m - 6n)$ continuous variables, and $(3mn - 2m - 2n + 1)$ unrestricted integer variables, not counting slack variables.

In the next section, we describe a new approach for improving the efficiency of the computation of the optimal bandwidth progressions in a network. This will make this method amenable for incorporation within a dynamic traffic management system.

5. A NETWORK DECOMPOSITION APPROACH

The principal difficulty in solving mixed-integer problems is the number of integer variables and their range, i.e., the size of the integer feasible set. Virtually all MILP codes use branch-and-bound strategies for calculating the optimal values. While there are now available quite efficient codes for solving large-scale problems, it is incumbent on all modelers, especially in traffic flow applications, to devise the most efficient solution procedure for a given problem. Clearly, more efficient computational procedures result in the ability to obtain improved solutions and, ultimately, lead to improved performance of the traffic network. Computational procedures are as important as are more accurate traffic models, especially in an era of increasing real time applications. For progression optimization in networks, this enables the following:

- the use of more economical mathematical programming codes which are more affordable and more easily accessible;
- the optimization of larger-scale networks than are otherwise possible;
- more reliable convergence to optimal solutions, i.e., the procedure does not fail as often;
- analysis of a larger number of alternatives at a reduced cost (the analysis of progression schemes involves a multitude of data sets, coefficients and parameters); and
- an opportunity to use the codes in real time, e.g., in a multilevel real time traffic-adaptive control system in conjunction with a dynamic traffic assignment and routing capability [10].

Branch-and-bound is essentially a strategy of "divide and conquer". The idea is to partition the feasible region into more manageable subdivisions and, eventually, to fathom the entire tree of integer solutions. However, branch-and-bound is a strategy for general purpose mixed-integer programs that does not exploit the special characteristics that a particular problem may have. Solving the primary multiband problem by general purpose branch-and-bound is still a rather formidable task. For this reason, heuristic methods which quickly lead to a good, though not always optimal solution are often preferable. Typically, heuristics have an intuitive justification motivated by an intimate familiarity with the particular problem characteristics [11].

It is virtually always advantageous to decompose mixed-integer programming problems into smaller subproblems, in order to reduce the number of integer variables that have to be considered in each subproblem, and to restrict as much as possible the allowable range of each variable. Several authors have proposed solutions for the uniform bandwidth problem along these lines. Mireault [12] tried to solve more efficiently an early version of the mixed integer linear programming progression model by carefully restricting the range of the integer variable set. Chaudhary

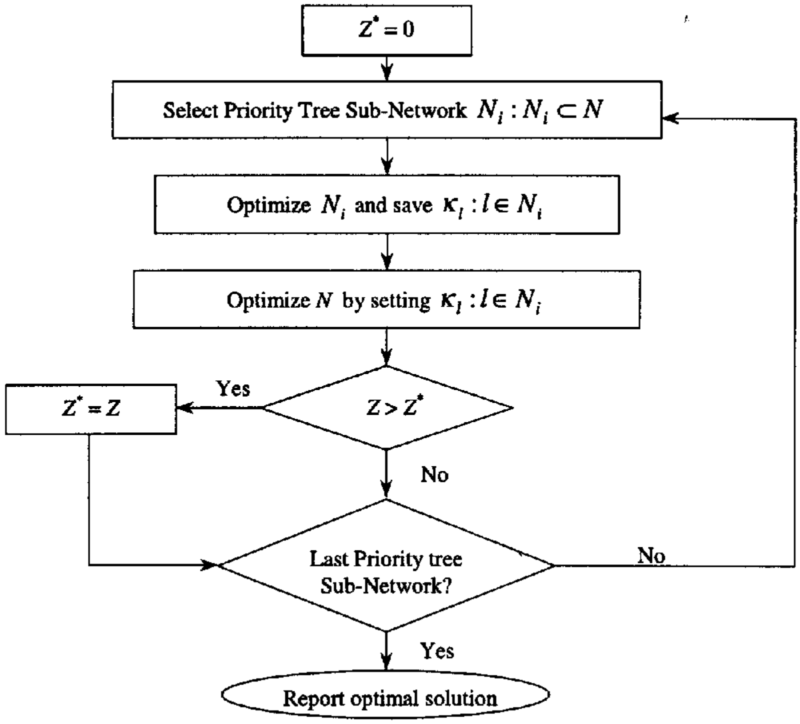

Figure 6. The network decomposition picture.

et al. [13] devised two decomposition procedures for the network version of MAXBAND by dividing the integer variable set into two or three subsets. Pillai *et al.* [14] proposed a two-part "greedy" heuristic which, similar to Chaudhary *et al.*, involves partial relaxation of some integer variables coupled with a depth-first search of the branch-and-bound tree. These approaches use an arbitrary fragmentation of the integer variable set and are not motivated by any traffic-based considerations. As a result of the relaxation of the integrality of some of the integer variables, any intermediate solutions to the MILP problem are not feasible from a traffic network standpoint.

The procedure that is described in this paper does not merely exploit the mathematical structure of the mixed-integer problem, as most other approaches do, but rather it is based on the traffic characteristics of the network. As such, it does not require any modification of the optimization code as is required by the other approaches. A *priority arterial subnetwork* consisting of an arterial tree is selected from the original traffic network based on its geometry and on the volumes of traffic flow that each link is carrying. The priority subnetwork is optimized first and the results are then used for the solution of the entire network. Alternative subnetworks can be selected in a heuristic procedure if further improvements are desired. Most importantly, this decomposition process does not require relaxation of any integer variables at intermediate steps. Thus, the intermediate solutions obtained at each step consist of a feasible and optimal solution for the priority subnetwork considered. This feature, in combination with the fact that the priority subnetwork contains the bulk of the traffic volumes in the network, enables the achievement of faster and better solutions. Computational results indicate that an optimal solution is obtained in the majority of cases during the first iteration. The decomposition procedure is described below.

Let $P1$ be a progression optimization problem obtained by considering only a subnetwork N_i of the original network N, and $P2$ be another progression optimization problem obtained by setting a selected subset of the arterial loop and the network loop integer variables in the network problem to a specific set of values. Then the heuristic network decomposition method, shown in Figure 6, is as follows.

- Step 1: Identify a new priority subnetwork $N_i \subset N$.
- Step 2: Optimize $P1$ for N_i, and save the resulting values of the arterial loop and network loop integer variables κ^*_{ij} for all the links loops of N_i.
- Step 3: Optimize $P2$ by setting the integer variables calculated in Step 2 ($\kappa_{ij} = \kappa^*_{ij} : \kappa_{ij} \in P1$).
- Step 4: Calculate the value of the objective function. If it is better than the previous solution, save it.
- Step 5: Stop if all priority subnetworks have been considered; otherwise go back to Step 1.

A priority subnetwork is chosen so that it contains only a "tree" of arterials, eliminating any network loops. The arterials contained in the "tree" should include the principal arterials of the network and can be chosen based on the following criteria.

1. Choose the principal arterial of the network to be the trunk of the tree and include only crossing arterials in the subnetwork.
2. The tree consists of the maximum number of arterial two-way links without forming any network loops.

The solution of both $P1$ and $P2$ can be obtained very quickly due to the reduced number of integer variables. Table 2 shows the number of integer variables in the primary problem and in

Table 2. Size of MILP-3 problem for an $m \times n$ closed grid network.

Variable	Uniform Band	Variable Band	Heuristic-Step1	Heuristic-Step2
$b, \bar{b}$	$2(m+1)$	$2(2mn-m-n)$	$2(mn-1)$	$2(2mn-m-n)$
z	1	1	1	1
$w, \bar{w}$	$2(2mn-m-n)$	$2(2mn-m-n)$	$2(mn-1)$	$2(2mn-m-n)$
v	$(2mn-m-n)$	$(2mn-m-n)$	$mn-1$	$mn-m-n+1$
μ	$(m-1)(n-1)$	$(m-1)(n-1)$	0	$(m-1)(n-1)$
Total Integers	$3mn-2m-2n+1$	$3mn-2m-2n+1$	$mn-1$	$2(mn-m-n)$
Example (no. of integers)				
4×6 network	53	53	23	30
3×7 network	44	44	20	24

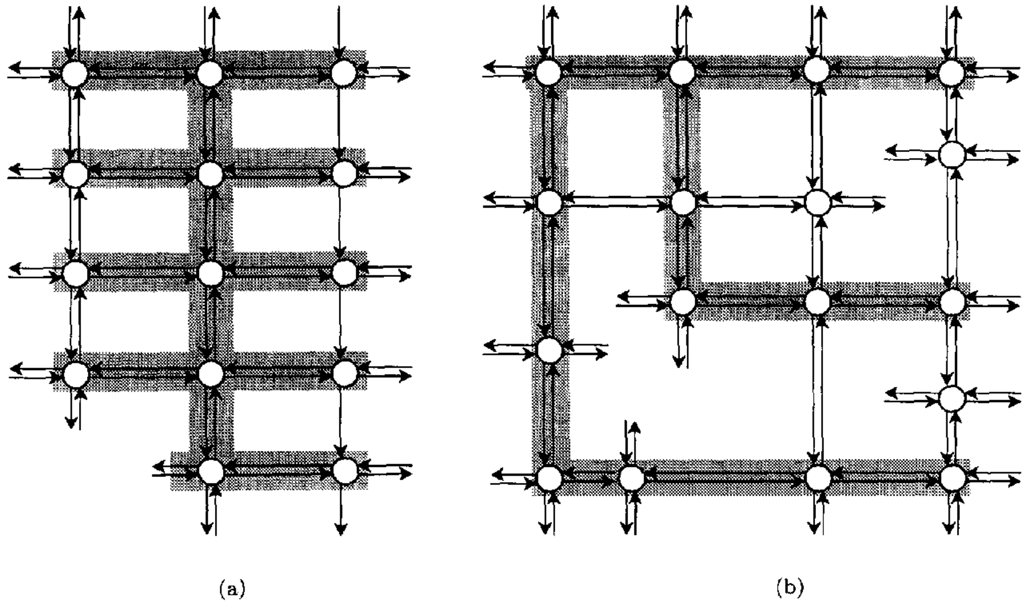

Figure 7. The (a) Ann Arbor, Michigan and the (b) Memphis, Tennessee networks.

the two subproblems $P1$ and $P2$ of the network decomposition approach for an $m \times n$ closed grid network ($m \times n$ intersections and $m + n$ arterials).

The performance of the heuristic solution was tested on two urban networks in the U.S.A.: one in Ann Arbor, Michigan (3×5 grid), shown in Figure 7a, and the other in Memphis, Tennessee (4×4 grid), shown in Figure 7b. The grids are not complete, as there are no signals at all nodes of the network. The shaded arterials indicate the priority arterial subnetworks utilized in this analysis. Computations were done on a 200 MHz Pentium PC. The procedure was applied to both the uniform and the variable bandwidth models. For the uniform bandwidth case, the heuristic calculated an optimal solution—there are multiple optimal solutions—for both sample networks. In case of the multiband model, the heuristic located the optimal solution only for one of the test networks. Execution times were dramatically reduced by as much as 1:263 compared to the original times (Table 3).

Table 3. Comparison of execution times for the primary and heuristic solutions.

Network/ Model	Size Art./Nodes	Primary MILP Problem Exec. Time	Heuristic Decomposition Exec. Time
Memphis, TN	8/17		
MILP-3 Uniform bandwidth		6,735 sec	50 sec (1/135)
MILP-3 Variable bandwidth		15,012 sec	57 sec (1/263)
Ann Arbor, MI	8/14		
MILP-3 Uniform bandwidth		4,235 sec	44 sec (1/96)
MILP-3 Variable bandwidth		9,995 sec	51 sec (1/196)

6. CONCLUSIONS

This paper presents mathematical programming models for the development of optimal arterial-based progression schemes in urban signal networks. Under such a scheme, a continuous green band is provided in each direction along the artery at the desired speed of travel to facilitate the movement of the principal through flows along the arterial. Both uniform- and variable-bandwidth models are formulated. The latter provide significant flexibilities for the development of optimal control strategies that are in harmony with the prevailing traffic flows in the network.

The paper also describes a heuristic network decomposition procedure for a more efficient solution of the mixed-integer linear programming formulation of the network progression problem. In contrast to previous approaches, which relax the integrality of the integer variables in the problem, this heuristic is based on the traffic characteristics of the network. The procedure produces intermediate solutions that are feasible and can be implemented directly in the network. It is shown that the procedure can yield considerable reductions in computational effort without or with little compromise in the quality of the results. Reductions range from 1:100 to 1:300 compared with the original formulation with the most dramatic reductions occurring in the case of the more complex problem of variable-bandwidth optimization.

By achieving these results, one can obtain more easily optimal solutions for large-scale networks, analyze a larger number of alternatives, as well as implement this strategy in a dynamic traffic management system. More efficient computational procedures result in improved solutions and, ultimately, lead to improved performance of the traffic network. They are as important as are more accurate modeling techniques.

REFERENCES

1. W.S. Homburger, Editor, *Transportation and Traffic Engineering Handbook*, Second Edition, Prentice-Hall, (1982).
2. R.L. Gordon *et al.*, *Traffic Control Systems Handbook*, Federal Highway Administration, Report FHWA-SA-95-032, U.S. Dept. of Transportation, Washington, DC, (1996).

3. M. Papageorgiou, Editor, *Concise Encyclopedia of Traffic & Transportation Systems*, Pergamon Press, (1991).
4. Forschungsgesellschaft für Strassen- und Verkehrswesen, *RiLSA: Richtlinien für Lichtsignalanlagen, Lichtzeichenanlagen für den Strassenverkehr*, Köln, Germany, (1992).
5. J.T. Morgan and J.D.C. Little, Synchronizing traffic signals for maximal bandwidth, *Operations Research* **12**, 896–912, (1964).
6. J.D.C. Little, The synchronization of traffic signals by mixed-integer linear programming, *Operations Research* **14**, 568–594, (1966).
7. J.D.C. Little, M.D. Kelson and N.H. Gartner, MAXBAND: A program for setting signals on arteries and triangular networks, *Transportation Research Record* **795**, 40–46, (1982).
8. N.H. Gartner, S.F. Assmann, F. Lasaga and D.L. Hou, A multi-band approach to arterial traffic signal optimization, *Transportation Research* **25B** (1), 55–74, (1991).
9. N.H. Gartner, Constraining relations among offsets in synchronized signal networks, *Transportation Science* **6**, 88–93, (1972).
10. N.H. Gartner and C. Stamatiadis, Integration of dynamic traffic assignment with real time traffic adaptive control, *Transportation Research Record* (1644, TRB), 150–156, (1998).
11. G.L. Nemhauser and L.A. Wolsley, *Integer and Combinatorial Optimization*, Wiley-Interscience, New York, (1988).
12. P. Mireault, A branch-and-bound algorithm for the traffic signal synchronization problem with variable speed, In *TRISTAN 1*, Montreal/Quebec City, (1991).
13. N.A. Chaudhary, A. Pinnoi and C. Messer, Proposed enhancements to MAXBAND-86 program, *Transportation Research Record* **1324**, 98–104, (1991).
14. R.S. Pillai, A.K. Rathi and S. Cohen, A restricted branch-and-bound approach for setting the left turn phase sequence in signalized networks, *Transportation Research* **32B** (5), (1994).