# Decision Model for Priority Control of Traffic Signals

Larry Head, Douglas Gettman, and Zhiping Wei

**This paper presents a model of the core logic of a traffic signal controller. The model is formulated on the basis of the traditional North American ring, phase, and barrier construct and includes phase intervals such as minimum and maximum times, pedestrian service, alternative minimum times, and a priority service extension. The mathematical model is based on precedence graphs that are familiar to engineers involved with project management techniques such as Gantt charts, the critical path method, and the program evaluation and review technique. The model presents an analytical framework for the analysis of complex controller behaviors and is demonstrated for the case of multiple priority requests. An example shows that a first-come, first-served policy for serving priority requests can result in more delay than will a multiple-priority-request policy generated by the model developed in this paper. Additional controller behaviors, such as preemption, coordination, and offset transition, can be analyzed with this model.**

Actuated signal control in North America has evolved steadily during the past 40 years. Today there are several standards for hardware interfaces (NEMA, CALTRANS 170/2070), new standards for communications (NTCIP), and emerging standards for traffic controller hardware and software (ATC). The evolution of the basic core logic of actuated signal controllers in North America has settled primarily on the concept of rings, barriers, and phases. Most commercially available traffic signal controllers are based on that construct with phase intervals used to provide safety constraints (minimum times), pedestrian service (walk, flashing don't walk, and don't walk), and maximum times. Coordination interacts with the core logic through hold and force-off commands that are determined by the coordination logic that establishes the cycle length, phase splits, and an offset.

As operating agencies have determined a need to extend this basic model of signal state transitions to address a specific need or to provide additional flexibility and functionality, controller manufacturers and firmware developers have responded by adding new features into their software. As a result, the controller state transition logic has become a complex collection of if–then rules that typically are described independently for each feature—such as first-come, first-served priority for transit vehicles. The effects of the interactions of multiple sets of rules that might become active for multiple demands on the controller (such as preemption and pedestrian service, or multiple transit priority requests) and the large number of parameter–value combinations that are possible make it challenging to understand and to predict controller behavior.

L. Head and Z. Wei, System and Industrial Engineering Department, University of Arizona, Tucson, AZ 85721. D. Gettman, Siemens Intelligent Transportation Systems, 6375 E. Tanque Verde #170, Tucson, AZ 85715.

*Transportation Research Record: Journal of the Transportation Research Board, No. 1978,* Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 169–177.

This paper presents a modeling framework for traffic signal control. The model presented represents the traditional core ring, barrier, and phase logic as a precedence graph, but is not limited to the ring, barrier, and phase construct. Representing the core logic as a precedence graph supports decision making for future controller states by considering the operational constraints of the traffic signal state transition logic. The precedence graph structure presented in this paper is analogous to the classical project management techniques of a Gantt chart, critical path method (cpm), and program evaluation and review technique (PERT) (*1*), but is formulated to address the structural and operational issues of traffic signal control directly. The goals for this new analytical model include providing the following:

- Structure for analyzing signal state transitions,
- Extensible framework to allow consideration of new features and functions, and
- More efficient signal timing that considers multiple objectives such as transit priority requests, vehicle demands, and pedestrian needs.

The objective of the precedence graph framework is to provide an accessible representation of traffic signal controller logic that provides a framework for analysis that is not readily available using nested sets of if–then rules. For example, with a nested rules approach, the validity of all the parameters is not necessarily checked (and in most cases cannot be checked) for all parameter range values. With the precedence graph representation, the validity of the network can be automatically verified by running a diagnostic procedure akin to cpm (*1*).

The second objective of this approach is to provide an extensible framework for considering advanced, alternative, and new features and functions. Within the structure, the approach is to specify the precedence structure to include different traffic phasing and layout relationships, additional operational constraints, different performance objectives, additional controller features or functionalities, or even different solution methods (i.e., optimization engines or heuristics).

Finally, the overarching objective of the precedence graph approach is to provide more efficient traffic signal timing through evaluation of alternative future timings given the wide range of phase parameters, phase flags, and special behaviors such as preemption and priority. The traditional if–then rules embedded in traffic signal controllers today can provide a vast array of behavioral characteristics, but must be set up precisely, by very skilled engineers, to take advantage of all of the available features and functions. Typically the operational transition from one timing state to another in the if–then framework is made on the basis of a limited set of events, such as gap-out, max-out, or force-off of the current phase and demand for the next phase in the sequence. In most situations, with appropriately configured splits, cycle, and offset values, coordinated-actuated operation provides fairly responsive signal timing. However, when anomalous conditions are experienced, such as preemption events, transit priority events, critical intersection conditions (oversaturated approaches), the if–then structures can be inefficient in responding to traffic conditions. The goal of the precedence graph modeling approach is to evaluate alternative timing schedules efficiently for the upcoming phases and priority requests that achieve or approach some operational goal, such as the following:

- Providing adequate pedestrian clearance time (or skipping pedestrian service) in preparation for a railroad preempt,
- Accommodating multiple transit priority requests by extending or truncating phases splits or resequencing phases,
- Transitioning back to the coordination point after a preemption event without causing oversaturation, and
- Extending a phase for dilemma zone protection for large vehicles such as trucks (*2*).

It is with these types of operational objectives in mind that the term "efficiency" is used in referring to the precedence structure approach instead of the term optimality, because optimality typically refers to minimization of traveler delay time or stops or both, which may not be the goal of the signal controller at a particular time. The precedence graph approach does not preclude objectives such as optimality of the signal timing, but it is not explicitly addressed in this paper.

## MODEL FORMULATION

### Modeling Ring, Barrier, Phase Constructs

For the purposes of this discussion assume a typical North American dual-ring, eight-phase controller (Figure 1). The approach presented here is completely extensible to controllers with more or fewer phases and rings and can accommodate any ring, barrier, and phase configuration. The precedence graph model is formed by representing each phase by an "activity on arc" task as shown in Figure 2. Figure 2 shows two cycles of the dual-ring controller.

There is no restriction to the ring, barrier, and phase construct, and it is possible to use the precedence graph to model compatible combinations of movements (signal indications).

Let $t_p^k$ represent the starting time of phase $p$ during the $k$th cycle and $v_p^k$ represent the duration of phase $p$ during the $k$th cycle. The phase duration $v_p^k$ includes the phase green time $g_p^k$ and the yellow ($y_p$) and all-red ($r_p$) clearance times, that is,

$$v_p^k = \begin{cases} g_p^k + y_p + r_p & \text{if } g_p^k > 0 \\ 0 & \text{if } g_p^k = 0 \end{cases} \tag{1}$$

where the yellow and red clearance times are phase input parameters and do not vary by cycle. It is possible that these intervals could be made variable on the basis of current state and conditions, such as the previous phase or a confliction movement. However, that behavior is not considered in this paper.

In Figure 2, the first cycle starts timing Phases 1 and 5. Phase 1 completes its duration after $v_1^1$ s, and then Phase 2 starts timing. Phase 5 times for $v_5^1$ s, and then Phase 6 starts timing. Phase 2 and 6 time for $v_2^1$ and $v_6^1$ s, respectively, but Phases 3 and 7 cannot start until both Phases 2 and 6 have completed timing. This is the traditional barrier crossing point of a dual-ring controller. Phases 1 and 5, in the second cycle, cannot start timing until after Phases 4 and 8 each complete timing during the first cycle. The precedence relationship requires this barrier crossing to be enforced.

The precedence graph can be represented as a set of precedence relationships assuming all timing starts at $t = 0$ and considers $K$ total cycles, and it can be written as follows:

$$\begin{aligned}
&t_1^1 = 0 \\
&t_5^1 = 0 \\
&t_2^k = t_1^k + v_1^k \\
&t_6^k = t_5^k + v_5^k \\
&t_3^k = t_2^k + v_2^k, \quad t_3^k = t_6^k + v_6^k \\
&t_7^k = t_2^k + v_2^k, \quad t_7^k = t_6^k + v_6^k \\
&t_4^k = t_3^k + v_3^k \\
&t_8^k = t_7^k + v_7^k \qquad \text{for } k = 1, \ldots, K \\
&t_1^{k+1} = t_4^k + v_4^k, \quad t_1^{k+1} = t_8^k + v_8^k \\
&t_5^{k+1} = t_4^k + v_4^k, \quad t_5^{k+1} = t_8^k + v_8^k \qquad \text{for } k = 1, \ldots, K-1
\end{aligned} \tag{2}$$

Consideration of multiple cycles allows modeling of the fact that the duration of a phase in one cycle can be different in subsequent cycles. When this precedence model is used to make decisions on servicing priority requests, the ability to consider multiple cycles and hence multiple service opportunities for priority service is important.

### Modeling Phase Interval Constraints

In the previous section the phase durations were considered as simple durations that might be constrained by simple phase minimum and phase maximum times. Each phase actually consists of a collection of phase intervals that determine the duration of the phase. Figure 3 illustrates one possible precedence graph representation of the phase intervals. The horizontal path through the graph represents the basic phase timing, which consists of the minimum green time (denoted $\text{min}_p$), the extension interval ($\text{ext}_p$), a priority extension interval ($\text{scp}_p$) (*3*), the yellow clearance ($y_p$), and the red clearance ($r_p$). The lower path represents the intervals required for servicing pedestrians including the walk interval ($w_p$), the pedestrian clearance or flashing don't walk ($\text{fdw}_p$), and the don't walk interval ($\text{dw}_p$). The upper path represents an alternative minimum ($\text{amin}_p$). The dashed arc represents a precedence relationship with no duration.

The alternative minimum time might be used in priority timing to constrain the phase minimum time when vehicle calls (or detector presence) are active. A typical application would be on a side-street left turn in which the phase might be forced off to provide an early green to another phase or in an emergency vehicle preemption situation in which it is desired to have additional green for the phase. The priority extension time is an interval that can be used to extend a phase to provide priority timing. Not all controllers use alternative minimum and priority extension intervals in that way, but the modeling approach presented here provides a construct to allow many alternative precedence relationships to be modeled for each phase.

These intervals can be divided into two general classes: fixed duration, which includes ($\text{min}_p$, $\text{amin}_p$, $\text{fdw}_p$, $y_p$, and $r_p$), and variable

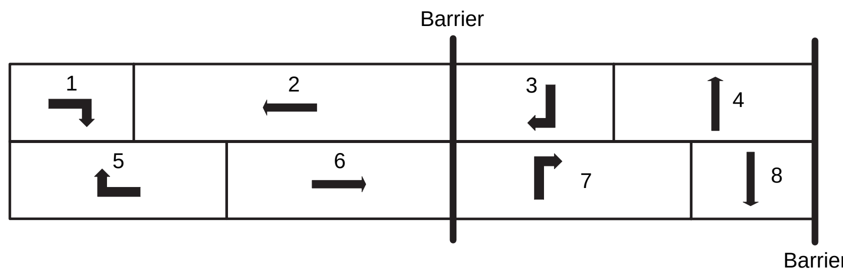



**FIGURE 1** Dual-ring, eight-phase controller.

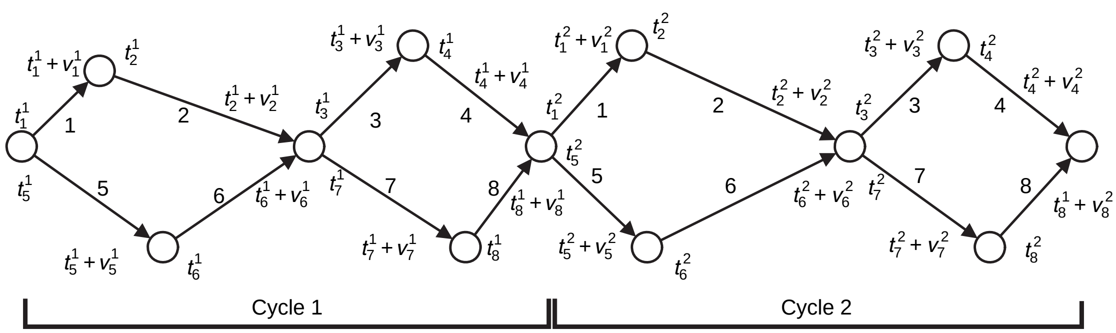



**FIGURE 2** Precedence graph representation of dual-ring controller.

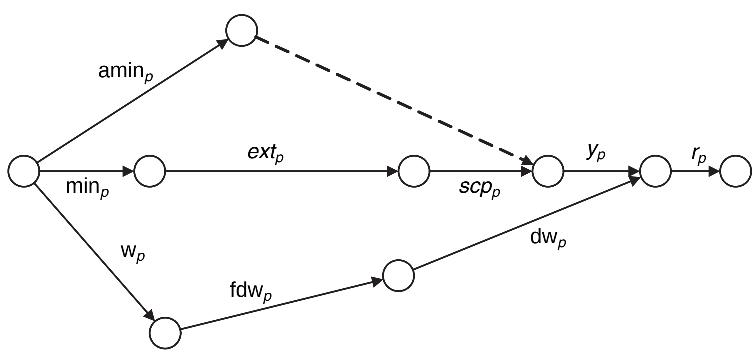



**FIGURE 3** Phase interval precedence diagram.

duration, which includes ($ext_p$, $cp_p$, and $w_p$, $dw_p$, or both). Fixed-duration intervals and the walk interval duration are provided as phase timing data. The variable duration intervals are determined by phase timing data (such as the vehicle extension interval), as part of the priority timing strategy (such as the priority extension interval), or to make sure that the interval path uses all available time (such as the walk, walk rest, and don't walk intervals). These variable duration intervals provide the flexibility needed to accommodate a wide variety of timing decisions. In the model presented in Figure 3, note that the pedestrian service interval ends at the end of the yellow clearance. This could be modeled to end at the end of the red clearance or at the beginning of the yellow clearance. Different controller manufacturers have implemented that differently, and some have provided the user with the option to select when this interval ends. For the purpose of this paper it is assumed to end at the end of the yellow clearance interval without any loss of generality.

Table 1 lists some of those timing interval parameters data and some of the phase flags. Each data value is named starting with a "*D*" to indicate data variable. Each flag is named starting with an "*X*" to indicate that it is a flag (on = 1, off = 0) variable. This notation will be used in the formulation below.

Only a subset of the parameters and flags are defined in the NTCIP 1202 actuated signal control specification (*4*) and NTCIP 1211 signal control and prioritization standard (*3*). They were selected because they capture the essence of controller behavior under most traffic applications when requests for priority are considered.

In addition to the phase parameters and flags, real-time data affect the phase durations. These variables are listed in Table 2. Each real-time call name is preceded by a "*C*" to indicate a call. These calls are in addition to the requests for priority.

The parameter data and flags are assumed to be structural parameters and do not change across cycles as phases are timing. Calls are real-time data and can arrive at any time during the phase timing process. Priority requests are assumed to be calls for service at some point in time (future) for a phase *p*. Extensions of the basic model to include priority requests are described below.

**TABLE 1 Phase Data Parameters**

| Phase Parameters (Data) | Name | Range |
|---|---|---|
| Minimum green | $D\mathrm{min}_p$ | 0–25.5 |
| Alternate minimum | $D\mathrm{amin}_p$ | 0–25.5 |
| Maximum green | $D\mathrm{max}_p$ | 0–255 |
| Walk | $Dw_p$ | 0–255 |
| Pedestrian clearance | $D\mathrm{fdw}_p$ | 0–255 |
| Extension | $D\mathrm{ext}_p$ | 0–25.5 |
| SCP extension | $D\mathrm{scp}_p$ | 0–255 |
| Yellow clearance | $Dy_p$ | 0–25.5 |
| Red clearance | $Dr_p$ | 0–25.5 |
| **Phase Flag Parameters** | **Name** | **Value** |
| Omit | $X\mathrm{omit}_p$ | {0,1} |
| Minimum recall | $X\mathrm{min}R_p$ | {0,1} |
| Maximum recall | $X\mathrm{max}R_p$ | {0,1} |
| Pedestrian recall | $X\mathrm{ped}R_p$ | {0,1} |
| Pedestrian omit | $X\mathrm{pedOmit}_p$ | {0,1} |
| SCP recall | $X\mathrm{scp}R_p$ | {0,1} |
| SCP omit | $X\mathrm{scpOmit}_p$ | {0,1} |

**TABLE 2 Real-Time Phase Interval Service Calls**

| Real-Time Calls | Name | Value |
|---|---|---|
| Phase call | $C\mathrm{phs}_p$ | {0,1} |
| Pedestrian call | $C\mathrm{ped}_p$ | {0,1} |
| Phase extension call | $C\mathrm{ext}_p$ | {0,1} |
| Priority request | $R_p^j$ | Time |

## Modeling Priority Requests

Priority requests are a special class of calls for service that originate from sources outside (e.g., from a priority request generator) the controller, are for a specific phase, and are for a time that is in the future (*3*, *5*). It is assumed that there can be one or more requests active at any time and that priority requests are not, by policy, handled on a first-come, first-served basis. Priority requests may come from transit systems, heavy rail, commercial vehicles, and adjacent intersections (for coordination purposes). It is also assumed that the prioritization of many requests is done externally to the controller and that only requests that have been selected for service are considered.

A priority request is denoted $R_p^j$ and represents a request for service for phase *p*. The superscript *j* denotes the *j*th request that is active for phase *p*. Figure 4 illustrates a single priority request for Phase 8 at time 42. Phase 8 does not start serving for 6 s (hence the request is labeled as having 6 s of delay). It is important to note that a set of requests is analogous to a set of phase or pedestrian calls, except that they are planned for service in the future and are not necessarily served until the associated requested phase is served. The problem of determining the phase durations to best serve the priority requests will be addressed as an application of the model below.

Given a controller structure represented as a precedence graph and set of precedence relations (Figure 2 and Equation 2), a set of phase parameters and flags (Table 1), and a set of real-time calls and priority requests (Table 2), the behavior of the core logic of the controller is essentially determined (that is not to say that the behavior is deterministic). A few details, such as the mapping of phase extension calls to the length of the phase extension interval, have been omitted, but the general controller model operation is complete.

## APPLICATION TO PRIORITY CONTROL DECISIONS

In this section the precedence-based controller logic model is applied to the problem of selecting phase durations that best serve multiple requests for priority. The goal will be to achieve the minimum delay for a set of several requesting vehicles (not all vehicles). Multiple priority requests can occur when several buses approach an intersection in a short period of time, or when more than one emergency vehicle is responding to a nearby incident, or any combination of these or other service requests.

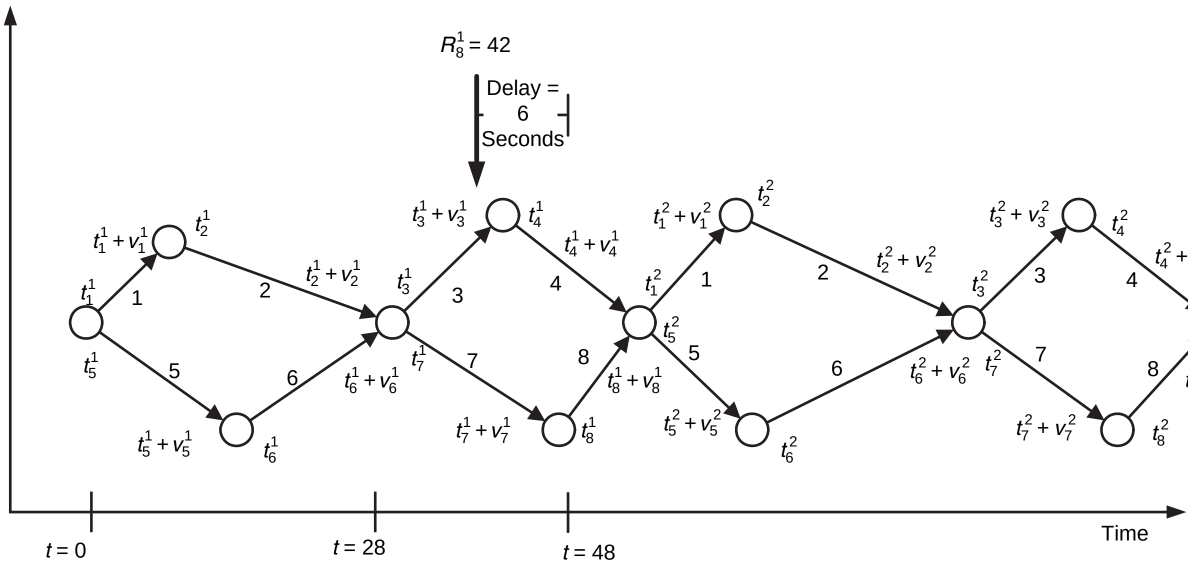



**FIGURE 4 Illustration of priority request at time = 42.**

The formulation consists of three parts:

1. Objective function to minimize the priority request delay,
2. Precedence relationship constraints that represent the controller phase and interval behaviors, and
3. Selection variables and constraints to determine the cycle containing service phase for each priority request.

The priority control problem is presented as a mathematical programming formulation as follows:

$$\begin{aligned}
&\text{minimize total priority delay}\\
&\text{s.t.}\\
&\quad\text{precedence constraints}\\
&\quad\text{phase duration constraints}\\
&\quad\text{service phase selection constraints}
\end{aligned} \tag{3}$$

The decision variables are the phase green times for each cycle, $g_p^k$, as well as decisions on whether to skip phases or phase intervals as needed to best serve the priority request. Table 3 lists the decision variables.

**TABLE 3 Decision Variables**

| Decision Variable | Name | Value |
|---|---|---|
| Phase green times | $g_p^k$ | $[0,\infty]$ |
| Phase interval start time | $t_p^k$ | $[0,\infty]$ |
| Skip ped. service | $Sped_p^k$ | $\{0,1\}$ |
| Skip minimum | $Smin_p^k$ | $\{0,1\}$ |
| Skip alt. minimum | $Samin_p^k$ | $\{0,1\}$ |
| Skip phase | $SC_p^k$ | $\{0,1\}$ |

It would be reasonable to assume that the traffic engineer might restrict some of the decision variables that allow a phase or a phase interval to be skipped. For example, it would be reasonable to allow skipping the pedestrian service intervals (walk, ped clearance, and don't walk), but not allow skipping the alternative minimum or the minimum time or skipping the phase. These restrictions would limit the ability of the controller to serve the priority requests satisfactorily, but would represent consideration of other factors, such as safety and quality of service for other intersection users.

The phase duration constraints for the problem presented here are simple nonnegativity constraints on the phase start times and range constraints on phase times. Hence,

$$\begin{aligned}
&t_p^k \geq 0,\\
&v_p^k = \begin{cases} g_p^k + y_p + r_p & \text{if } g_p^k > 0 \\ 0 & \text{if } g_p^k = 0 \end{cases}\\
&\min g_p \leq g_p^k \leq \max g_p \quad \text{for all } p \text{ and } k
\end{aligned} \tag{4}$$

The phase green minimum, $\min g_p$, and maximum times, $\max g_p$, are determined by the phase data parameters, flags, and real-time calls as follows:

$$\min g_p = \max \begin{bmatrix} D\text{min}_p, \\ D\text{amin}_p \bullet C\text{phs}_p \left(1 - S\text{amin}_p^k\right), \\ \left(Dw_p + D\text{fdw}_p\right) \bullet C\text{ped}_p \bullet \left(1 - S\text{ped}_p^k\right), \\ D\text{max}_p \bullet X\text{max}R_p \end{bmatrix} \bullet \left(1 - SC_p^k\right) \bullet \left(1 - X\text{omit}_p\right) \tag{5}$$

$$\max g_p = \left[\max\left(D\text{max}_p, \left(Dw_p + D\text{fdw}_p\right) \bullet C\text{ped}_p\right)\right] \bullet \left(1 - SC_p^k\right) \bullet \left(1 - X\text{omit}_p\right) \tag{6}$$

Equations 5 and 6 essentially represent the shortest path and longest path for one particular phase in the precedence model.

The phase service selection constraints address the need to determine which phase service occurrence during the next few cycles is assigned to provide the service to each priority request. Figure 5 depicts the decision variables used to make that selection. Figure 5 shows three candidate service intervals (i.e., instances of the phase being served) in a three-cycle problem for a priority request for phase $p$: one service opportunity from $t_p^1$ to $t_p^1 + g_p^1$, one from $t_p^2$ to $t_p^2 + g_p^2$, and one from $t_p^3$ to $t_p^3 + g_p^3$. The $j$th priority request for phase $p$ occurs at time $R_p^j$. To determine which interval the priority request is being assigned to for service, two binary indicator variables, $\theta_{p,e}^{j,k}$ for the period before the service phase and $\theta_{p,s}^{j,k}$ for the period during the service phase for the $j$th cycle, are defined. For example, in Figure 5 the priority request occurs before the second candidate service intervals, hence it will be delayed before being served. In this case the service interval selection variable $\theta_{p,e}^{j,2} = 1$, and all others are zero.

The phase service selection constraints are modeled so that only one indicator variable—$\theta_{p,e}^{j,k}$ or $\theta_{p,s}^{j,k}$—is nonzero at any time. The phase service selection constraints for a three-cycle problem are as follows:

$$\begin{aligned}
t_p^1 - R_p^j &\geq \left(\theta_{p,e}^{j,1} - 1\right)M \\
t_p^1 + g_p^1 - R_p^j &\geq \left(\theta_{p,e}^{j,1} + \theta_{p,s}^{j,1} - 1\right)M \\
t_p^2 - R_p^j &\geq \left(\theta_{p,e}^{j,1} + \theta_{p,s}^{j,1} + \theta_{p,e}^{j,2} - 1\right)M \\
t_p^2 + g_p^2 - R_p^j &\geq \left(\theta_{p,e}^{j,1} + \theta_{p,s}^{j,1} + \theta_{p,e}^{j,2} + \theta_{p,s}^{j,2} - 1\right)M \\
t_p^3 - R_p^j &\geq \left(\theta_{p,e}^{j,1} + \theta_{p,s}^{j,1} + \theta_{p,e}^{j,2} + \theta_{p,s}^{j,2} + \theta_{p,e}^{j,3} - 1\right)M \\
t_p^3 + g_p^3 - R_p^j &\geq \left(\theta_{p,e}^{j,1} + \theta_{p,s}^{j,1} + \theta_{p,e}^{j,2} + \theta_{p,s}^{j,2} + \theta_{p,e}^{j,3} + \theta_{p,s}^{j,3} - 1\right)M
\end{aligned} \tag{7}$$

and

$$\begin{aligned}
R_p^j - t_p^1 &\geq -\theta_{p,e}^{j,1} M \\
R_p^j - \left(t_p^1 + g_p^1\right) &\geq -\left(\theta_{p,e}^{j,1} + \theta_{p,s}^{j,1}\right)M \\
R_p^j - t_p^2 &\geq -\left(\theta_{p,e}^{j,1} + \theta_{p,s}^{j,1} + \theta_{p,e}^{j,2}\right)M \\
R_p^j - \left(t_p^2 + g_p^2\right) &\geq -\left(\theta_{p,e}^{j,1} + \theta_{p,s}^{j,1} + \theta_{p,e}^{j,2} + \theta_{p,s}^{j,2}\right)M \\
R_p^j - t_p^3 &\geq -\left(\theta_{p,e}^{j,1} + \theta_{p,s}^{j,1} + \theta_{p,e}^{j,2} + \theta_{p,s}^{j,2} + \theta_{p,e}^{j,3}\right)M \\
R_p^j - \left(t_p^3 + g_p^3\right) &\geq -\left(\theta_{p,e}^{j,1} + \theta_{p,s}^{j,1} + \theta_{p,e}^{j,2} + \theta_{p,s}^{j,2} + \theta_{p,e}^{j,3} + \theta_{p,s}^{j,3}\right)M
\end{aligned} \tag{8}$$

where

$$\begin{aligned}
&\sum_k \theta_{p,e}^{j,k} + \theta_{p,s}^{j,k} = 1 \\
&\qquad \theta_{p,e}^{j,k} \in \{0,1\} \\
&\qquad \theta_{p,s}^{j,k} \in \{0,1\}
\end{aligned} \tag{9}$$

$M$ = number that is larger than the length of the service horizon.

These constraints act by selecting the time interval in which the request can be served given the current values of the decision variables.

Given the phase service selection variables, it should be noted that delay occurs only in the "early" arrival intervals; hence the objective function to minimize delay to the priority request is as follows:

$$\text{minimize} \quad D = \sum_{(j,p)} \sum_k \theta_{p,e}^{j,k} \left(t_p^k - R_p^j\right) \tag{10}$$

where the first summation is for every priority request $j$ and the requested phase $p$.

The priority control decision problem is a mixed-integer mathematical programming problem that can be solved by using readily available tools. The authors have used the Excel Solver and the LINGO Excel extension to solve problems consisting of several cycles and many priority requests in reasonable solution times.

## EXAMPLE—MULTIPLE TRANSIT PRIORITY REQUESTS

This example demonstrates how the priority controller formulation can be used to serve multiple priority requests. Traditionally, a first-come, first-served policy has been used in serving priority requests. Although that might yield a reasonable solution, it is not difficult to conceptualize a scenario in which the delay resulting from a first-come, first-served policy will result in more total delay to the vehicles (e.g., buses) receiving priority than would result if no attempt to provide priority were made at all. In this example the ability to skip phases and phase intervals is not considered as is reasonable for providing priority for transit vehicles and other "lower-priority" vehicles.

Table 4 contains the signal timing data, phase flags, and real-time calls for this example. All values are reported, but some features, such as the SCP extension capability, are not used in this example.

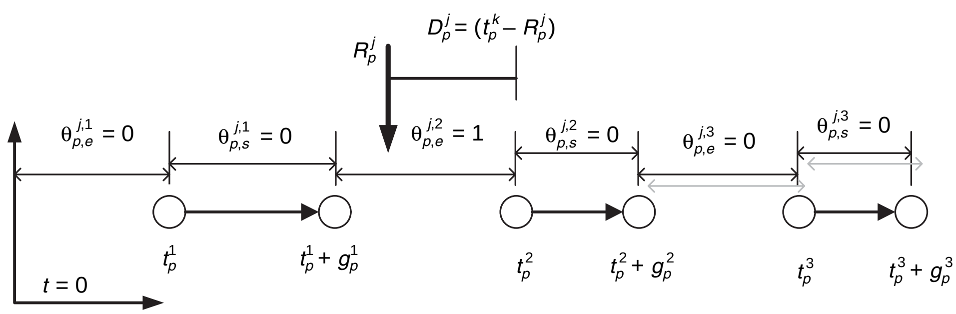



**FIGURE 5 Phase service selection variable definitions.**

**TABLE 4 Signal Timing Data, Phase Flags, and Real-Time Calls**

| | Name/Phase | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| Phase parameters (data) | | | | | | | | | |
| Minimum green | $D\mathrm{min}_p$ | 7 | 12 | 5 | 15 | 7 | 12 | 5 | 15 |
| Alternate minimum | $D\mathrm{amin}_p$ | 7 | | 5 | 15 | 7 | | 5 | 15 |
| Maximum green | $D\mathrm{max}_p$ | 18 | 55 | 23 | 45 | 18 | 55 | 23 | 45 |
| Walk | $Dw_p$ | | 7 | | 4 | | 7 | | 4 |
| Pedestrian clearance | $D\mathrm{fdw}_p$ | | 21 | | 23 | | 21 | | 23 |
| Extension | $D\mathrm{ext}_p$ | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| SCP extension | $D\mathrm{scp}_p$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Yellow clearance | $Dy_p$ | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| Red clearance | $Dr_p$ | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| Phase flag parameters | | | | | | | | | |
| Omit | $X\mathrm{omit}_p$ | — | — | — | — | — | — | — | — |
| Minimum recall | $X\mathrm{min}R_p$ | x | x | x | x | x | x | x | x |
| Maximum recall | $X\mathrm{max}R_p$ | — | — | — | — | — | — | — | — |
| Pedestrian recall | $X\mathrm{ped}R_p$ | — | — | — | — | — | — | — | — |
| Pedestrian omit | $X\mathrm{pedOmit}_p$ | x | — | x | — | x | — | x | — |
| SCP recall | $X\mathrm{scp}R_p$ | — | — | — | — | — | — | — | — |
| SCP omit | $X\mathrm{scpOmit}_p$ | x | x | x | x | x | x | x | x |
| Real-time calls | | | | | | | | | |
| Phase call | $C\mathrm{phs}_p$ | x | x | x | x | x | x | x | x |
| Pedestrian call | $C\mathrm{ped}_p$ | — | x | — | — | — | — | — | — |
| Phase extension call | $C\mathrm{ext}_p$ | — | — | — | — | — | — | — | — |

All phases have minimum recall set to TRUE (x) so that each phase will be required to serve. That is not necessary in general. It is important to note that the real-time phase calls represent a snapshot of the call state at the start of the priority timing interval. These values will affect the behavior of the first cycle, but their effect on future cycles is unknown because their values are not known.

Assume initially that the signal had been timing Phases 2 and 6 for 5 s when a priority request for Phase 8 is received. This priority request ($R_8^1$) is for Time 42. Figure 6*a* illustrates the time that the priority request is received. Notice that the precedence graph starts from a dummy node at time $t_0$ and represents the fact that Phases 2 and 6 have timed for 5 s when the priority timing problem is to be solved. Figure 6*b* shows the timing modified after solving the priority timing problem. There is now no delay to the priority request. This is essentially an early green for Phase 8.

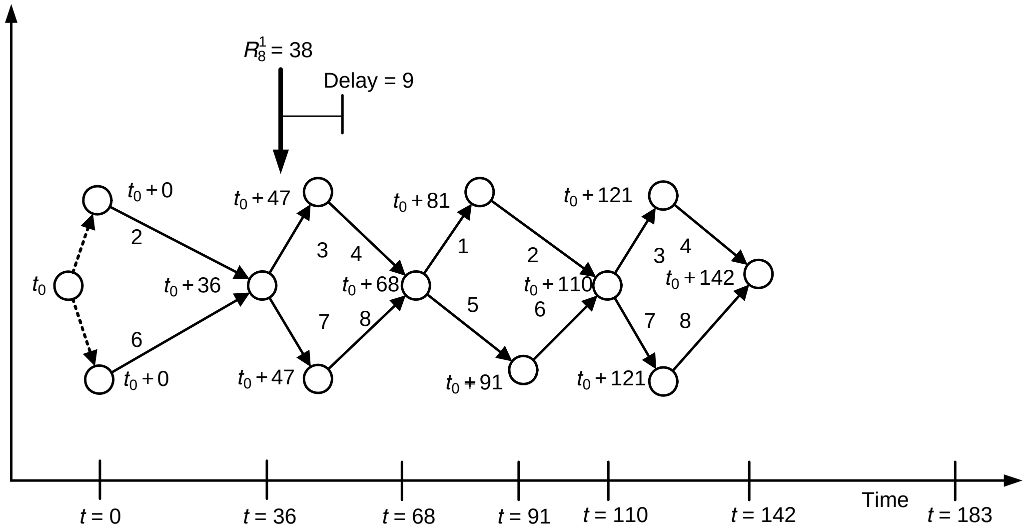
Now consider the receipt of a second request after 2 additional seconds as shown in Figure 7*a*. Assume that this second request is for Phase 2 at Time 41 ($R_2^1$). If the requests are served on a first-come, first-served basis the second request will incur 31 s of delay before Phase 2 is served. If both requests are considered using the precedence graph model and solving the associated optimization problem stated in Equation 3, then the early green decision for Phase 8 is changed to be a green extension for Phase 2 to serve the second request. This decision will delay the request for Phase 8 by 20 s, which is worse than when it was initially received, but results in less total delay for both requests. The total delay for both requests is reduced from 31 s to 20. Figure 7*b* shows the final signal timing and delay to the two priority requests.

This example illustrates how traditional first-come, first-served logic can actually result in more delay to vehicles requesting priority than can be achieved by using the decision model that considers all priority requests when selecting the signal timing.

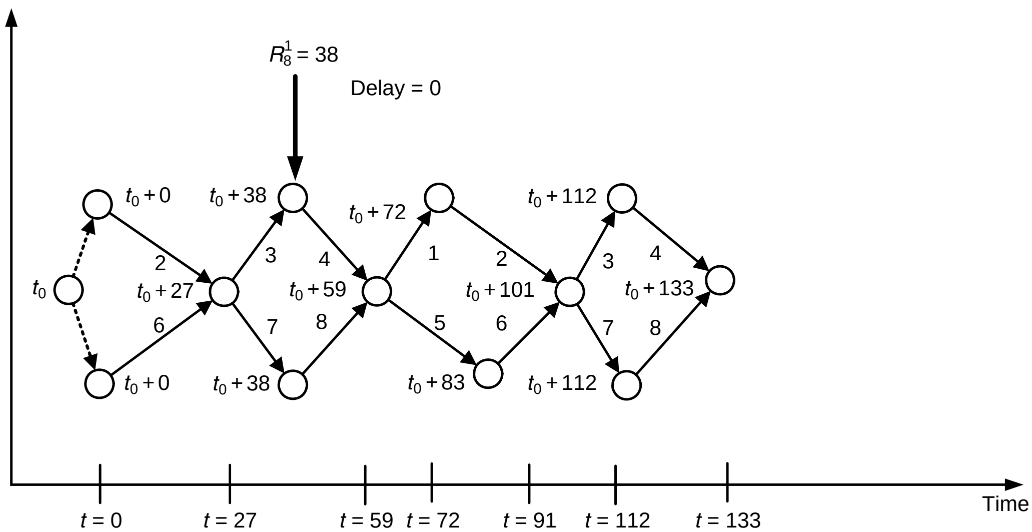
## DISCUSSION AND CONCLUSION

This paper presents a model that captures the core logic of a traffic signal controller in an analytical framework that can be used for simultaneous analysis of the behavior of ring, barrier, and phase based controllers as well as the intricate phase interval timing. The model is applied to show how multiple priority requests can be served efficiently while considering all of the constraints posed by phase timing data and flag parameters, as well as other real-time phase calls. The example presented is relatively simple, but illustrates the potential benefit in developing strategies that are not simply first-come, first-served in which priority requests can be received with sufficient lead time to allow intelligent service planning.

The analytical framework presented lends itself to analysis of many different controller behaviors, including preemption, coordination, transition, and priority (as exemplified). Other behaviors, such as phase reservice, phase sequence rotations (e.g., lead–lag to lag–lead), and overlaps, need to be considered in this framework. The authors believe that all of these behaviors can be modeled appropriately. The model can be expanded to more complex ring structures including three-phase and four-phase diamonds, multiple intersection control–coordination, and those that current controller




(a)




(b)

**FIGURE 6** Example signal timing (*a*) when first priority request is received and (*b*) after solving priority timing problem for Phase 8.

manufacturers use (e.g., Econolite: two-ring, 12-phase; Eagle: four-ring, 16-phase; and Siemens NextPhase: 20-ring, 40-phase). The model can also be used to capture allowable combinations of movements and phase intervals at complex intersections in which the ring, barrier, phase construct can be limiting.

The issue of computational complexity remains to be fully addressed. The current priority control formulation is a mixed-integer mathematical programming formulation. Past experience in the traffic control field (*6, 7*) has raised this as a major concern; but there have been significant advances in the mathematical programming science in the past 30 years, and in today's operations research standards, the size and complexity of the problem presented here are considered trivial. Libraries such as CPLEX and LINGO can easily solve these problems in a fraction of a second. In addition to the power of modern optimization libraries, formulations and solution using other optimization approaches including dynamic programming have been shown to be implementable (*8, 9*) on field-deployable hardware. The authors see no insurmountable barriers to implementation on modern traffic controller processors.

The model presented in this paper can be used to better understand the complex behavior of modern traffic signal control devices. It provides an analytical framework that includes all of the structures of rings, barriers, phases, and phase intervals. It offers a model for analyzing behaviors that need to be better understood to improve the efficiency of traffic signal operations.

## ACKNOWLEDGMENTS

This work has been performed under the NCHRP 3-66 Project, Traffic Signal State Transition Logic Using Enhanced Sensor Information. The authors acknowledge several individuals who have shared ideas and thoughts on the material presented in this paper: Tom Urbanik, Darcy Bullock, Pitu Mirchandani, and Steve Shelby.

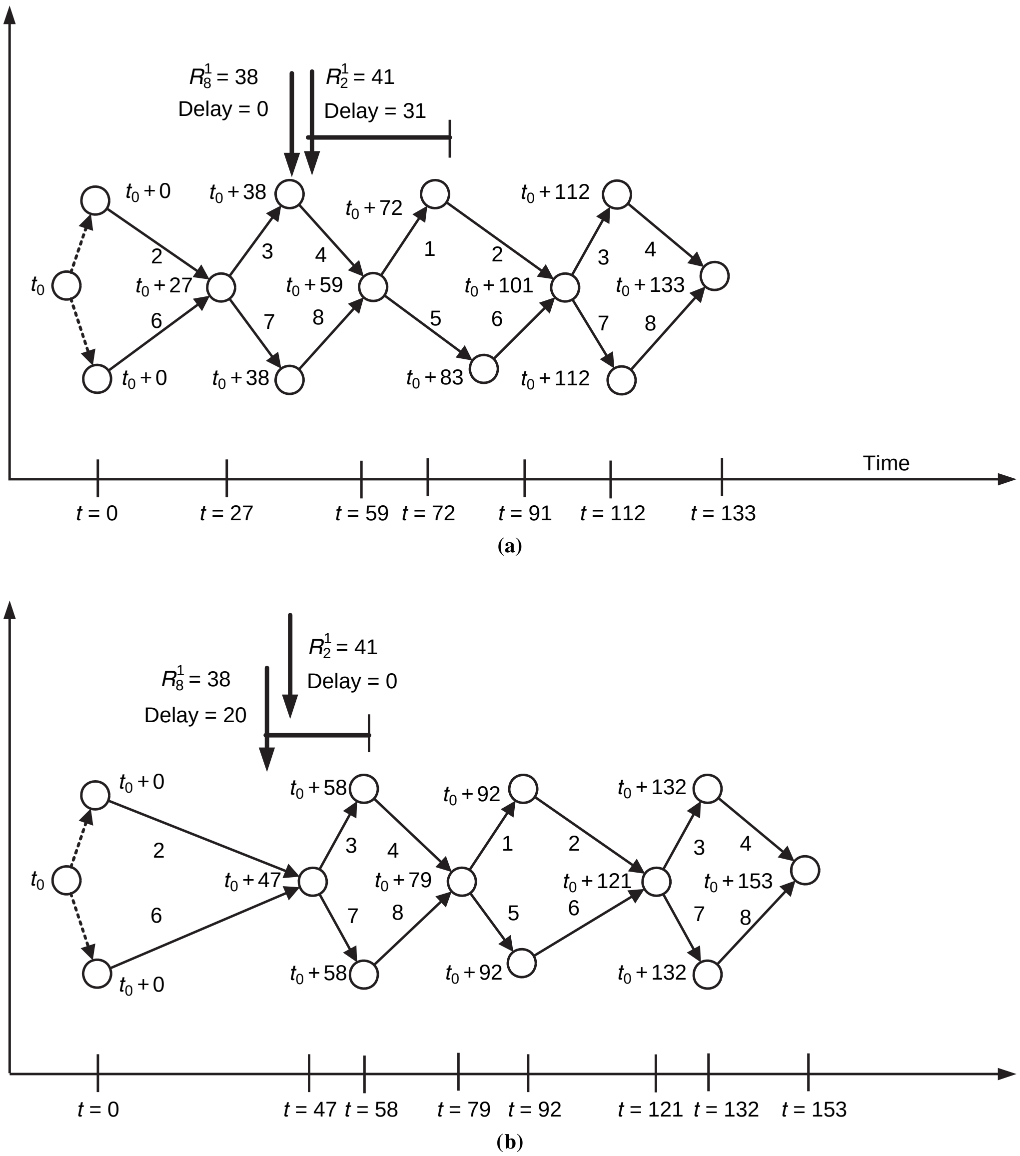



**FIGURE 7 Example signal timing after providing priority for first request and receiving a second request that is served (*a*) on a first-come, first-served basis and (*b*) after resolving the priority timing that considers both requests.**

## REFERENCES


1. Molder, J. J., C. R. Phillips, and E. W. Davis. *Project Management with CPM, PERT, and Precedence Diagramming.* Blitz Publishing, Middleton, Wis., 1995.
2. Sunkari, S. R., H. Charara, and T. Urbanik II. Minimizing Truck Stops at High-Speed Rural Signalized Intersections. *2001 Annual Meeting and Exhibit of the Institute of Transportation Engineers* (CD-ROM), ITE, Chicago, Ill., 2001.
3. *NTCIP 1211: Object Definitions for Signal Control and Prioritization.* National Electrical Manufacturers Association, Rosslyn, Va., 2004.
4. *NTCIP 1202: Object Definitions for Actuated Traffic Signal Controller Units.* National Electrical Manufacturers Association, Rosslyn, Va., 2004.
5. Head, K. L. *TCRP A-16 Interim Report: Improved Traffic Signal Priority for Transit.* Transit Cooperative Research Program, TRB, 1999. www4.nas.edu/trb/crp.nsf/All+Projects/TCRP+A-16A.
6. Gartner, N. H. Optimization of Traffic Signal Settings by Mixed-Integer Linear Programming, II: The Network Synchronization Problem. *Transportation Science,* Vol. 9, No. 4, 1975, pp. 344–363.
7. Gartner, N. H., F. J. Pooran, and C. M. Andrews. Optimized Policies for Adaptive Control Strategy in Real-Time Traffic Adaptive Control Systems: Implementation and Field Testing. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1811,* Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 148–156.
8. Sen, S., and K. L. Head. Controlled Optimization of Phases (COP) at an Intersection. *Transportation Science,* Vol. 1, No. 31, 1997, pp. 5–17.
9. Head, K. L., P. B. Mirchandani, and D. Sheppard. Hierarchical Framework for Real-Time Traffic Control. In *Transportation Research Record 1360*, TRB, National Research Council, Washington, D.C., 1992, pp. 82–88.


*The material presented in this paper is the authors' and is not necessarily that of the Transportation Research Board or NCHRP.*

*The Traffic Signal Systems Committee sponsored publication of this paper.*