Paper No. 02-3261

Dynamic Signal Coordination for Networks with Oversaturated Intersections

Montty Girianna and Rahim F. Benekohal

An algorithm to design signal coordination for networks with oversaturated intersections is presented. The basic concept of signal coordination applied to oversaturated single arterials is extended for a grid network of arterials, which involves greater analytical and computing complexity. In this algorithm, signal coordination is formulated as a dynamic optimization problem. The problem is developed to coordinate oversaturated signals along an arterial that crosses multiple, parallel coordinated arterials. Signals along crossing arterials are also oversaturated. During an oversaturated period, the algorithm manages local queues by spatially distributing them over a number of signalized intersections and by temporarily spreading them over signal cycles. Depending on the traffic demand's variation and the position of critical signals, the algorithm intelligently generates optimal signal timing (green times and offsets) along individual arterials. If critical signals are located at the exit points, the algorithm sets the optimal signal timing that protects them from becoming excessively loaded. If critical signals are located at the entry points, the algorithm ensures that queues are reduced or cleared before released platoons arrive at a downstream signal system. In addition, the algorithm eventually finds a set of common cycles propagated from upstream signals, thus promoting traffic progression. The micro-genetic algorithm was used to solve the signal optimization problem. The algorithm was tested on a one-way arterial system with 20 signals. The results indicate that the algorithm successfully managed queues along coordinated arterials, made the signals share the burden of traffic, and created the opportunity for traffic progression in specified directions.

Most existing signal coordination algorithms for intersection networks do not directly consider the dynamic evolution of queues at intersection approaches, so their application to oversaturated conditions may lead to suboptimal results. Oversaturation occurs when a signalized intersection cannot process all arrived vehicles at the end of a green period; thus, a queue is developed and carried over to the next green period. If corrective steps are not taken, the growing queue will block the intersection and reduce the capacity of intersection networks. Two widely used softwares, TRANSYT and PASSER, are valid only for undersaturated conditions. Their fixed signal timings are based on average situations and, therefore, do not consider the dynamic evolution of traffic at intersection approaches. Moreover, SCOOT uses a real-time scheme, but it utilizes a traffic flow model that does not include explicit oversaturation phenomena, such as blockage, carryover queue, and so forth.

Department of Civil and Environmental Engineering, University of Illinois at Urbana–Champaign, Newmark Civil Engineering Laboratory, 205 North Mathews Avenue, Urbana, IL 61801.

The work by Abu-Lebdeh and Benekohal (*1*, *2*) and Lieberman et al. (*3*) provides frameworks for developing a signal coordination model on arterials with oversaturated intersections. The works are based on dynamic queue management of a signal system on a single arterial. In this paper, the queue management concept is expanded to a network with oversaturated intersections. The basic philosophy is to ensure that signals share the burden of excessive traffic load during an oversaturation period, while simultaneously providing the opportunity for traffic progression for a certain direction. At the beginning of an oversaturated period, signal timings are set for queues to be distributed among signals along arterials. At this stage, removing the queues is the prime objective. Queue growth and dissipation are controlled to prevent intersection blockages. Once the queues are reduced or cleared, the next step is to provide a set of signal timing for traffic progression. It is obvious that this signal coordination philosophy requires algorithms that can work appropriately in both oversaturated and undersaturated conditions.

In this study, first the signals along a single arterial are coordinated; then, signals along the crossing arterials are coordinated. The first level of coordination is accomplished by changing offsets and green times along that arterial. Shorter green times at upstream signals and negative offsets are used at the beginning of an oversaturated period to allow sufficient times for the downstream signals to reduce or clear the queues before the arrival of new vehicles. During this period, the downstream signals must turn green before the upstream signals, which results in a green wave that moves upstream toward the drivers in the platoon (reverse progression). Then, forward progression is developed after the queues are cleared from those signals. This signal coordination along the arterial is bounded by critical signals, which are the intersections between two coordinated arterials.

Because the signal coordination algorithm finds optimal signal timing for the entire period of oversaturation, the problem of signal coordination becomes a large combinatorial optimization problem. This paper uses the micro-genetic algorithm (μGA) to solve the problem. GAs are search methods based on the mechanics of natural selection and natural genetics (*4*). They are regarded as emerging optimization tools because of their simplicity, minimal problem restrictions, capability of solving large combinatorial problems, parallelism, and productivity of global solutions. GAs simulate the genetic state or chromosome of an individual population by selecting the most influential individual through genetic operators such as natural selection, mutation, and crossover. For signal coordination problems, the chromosome represents the set of green times for the entire period of oversaturation.

This paper presents signal coordination for a network with oversaturated intersections formulated as a dynamic optimization problem. It describes a GA-based procedure used to solve the optimization problem and demonstrates application of the procedure on an arterial network with 20 intersections.

FORMULATION OF SIGNAL COORDINATION

This section describes a model of signal coordination problems for oversaturated networks called the base signal network problem (BSNP). Signals are coordinated along a single arterial that crosses coordinated parallel arterials. All intersections are oversaturated. By definition, BSNP does not coordinate signals around a loop.

The formulation of BSNPs presented here is for a two-phase signal plan without turning movements. Let $G = (N, L)$ denote a traffic signal network consisting of a set of signals, N, and a set of directional (one-way) streets, L. For a one-way arterial network, as indicated in Figure 1*a*, $N = 20$ and $L = 49$ (one-way including exit and entry links). Suppose that, during an oversaturated period, one tries to coordinate the signals along a northbound arterial and those along arterials running east–west, indicated by directed thick lines. Denote a set of coordinated signals from signal i to signal j as p_{ij}. For the signal network presented in the figure, $p_{1,5} = \{1, 2, 3, 4, 5\}$, $p_{10,6} = \{10, 9, 8, 7, 6\}$, $p_{11,15} = \{11, 12, 13, 14, 15\}$, $p_{20,16} = \{20, 19, 18, 17, 16\}$, and $p_{20,5} = \{20, 15, 10, 5\}$. Signals along other paths in the network are noncoordinated.

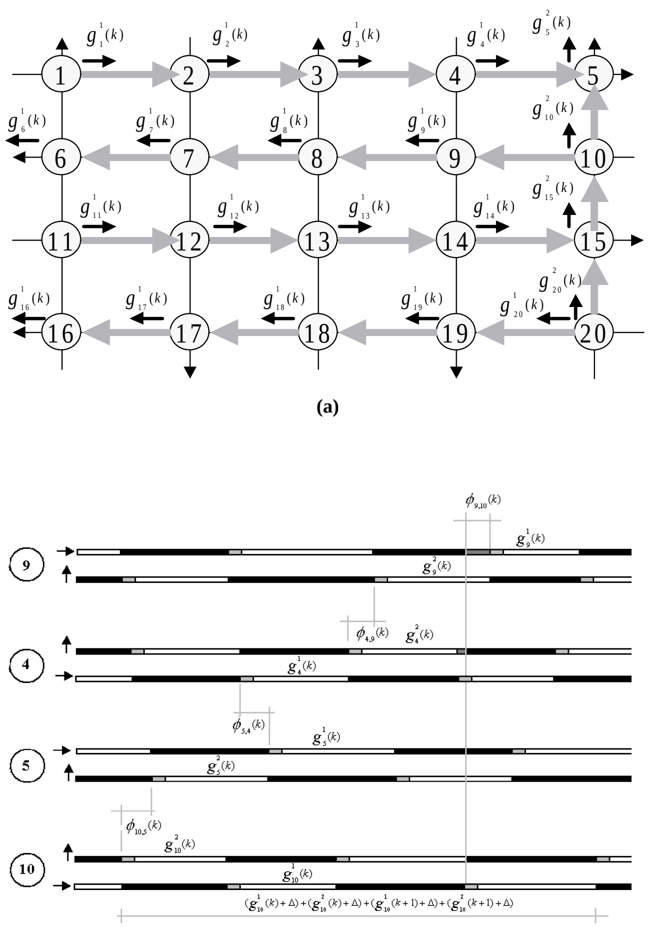


FIGURE 1 *(a)* **One-way arterial network with 20 signals: thick arrows indicate coordinated arterials.** *(b)* **Space–time diagram for Signal Loop 10–5–4–9.**

Objective Functions

Let Z be a network performance function as formulated by Equation 1. This function represents the number of vehicles processed by a signal network minus a network disutility function, which represents the occurrence of a queue on signal approaches along coordinated arterials. The overall goal is to maximize the net effect. K is the period of oversaturation in a cycle unit, and $D_{i,j}(k)$ is the departure flow at signal j coming from signal i at cycle k. A nonnegative weighting factor, $l_{i,j}$, is the distance between signal i and j, and $q_{i,j}(k)$ is the number of vehicles in the queue approaching signal j coming from signal i at the beginning of cycle k. L_p is a set of streets carrying coordinated movements, and $\delta_{i,j}(k)$ is a nonnegative disutility factor whose values are determined based on a queue management strategy. In this paper, $\delta_{i,j}(k) = 1$ for $(i, j) \in L_p$ and $\delta_{i,j}(k) = 0$ otherwise.

$$\max z = \sum_{k}^{K} \sum_{i,j \in L} l_{i,j} D_{i,j}(k) - \sum_{k}^{K} \sum_{(i,j) \in L_p} \delta_{i,j}(k) q_{i,j}(k) \qquad \delta_{i,j}(k) > 0, l_{i,j} > 0 \tag{1}$$

Offset and Green Time Constraints: Coordinated Arterials

To provide traffic progression, two constraints need to be satisfied: efficient use of green time and de facto red avoidance (*3*). For efficient use of green time, an offset between two adjacent signals must be assigned so that the queue at the approach of downstream signals is cleared and the first vehicle released from upstream signals reaches the tail of the (moving) queue. This offset is called the ideal offset. For a two-signal system (signals i and j), as indicated in Figure 2*a*, the time required for the tail to start moving, $\alpha_{i,j}(k)$, is calculated by dividing the queue length at the approach of signal j, $q_{i,j}(k)$, by the starting shock-wave speed. $\tau_{i,j}(k)$ is the time required for the first vehicle in the released platoon from signal i to join the tail of the downstream platoon (as the tail has reached its desired speed). It is determined by the acceleration rate, the existence of queues at the approach of signal i, and the availability of an unoccupied space along the street connecting the two signals. Thus, for example, when the unoccupied space is long enough that it provides the platoon's first vehicle a sufficient distance to reach a free-flow speed, $\tau_{i,j}(k)$ is simply the ratio of the space to the free-flow speed. The ideal offset between signals i and j, $\phi_{i,j}(k)$, is set equal to the difference between $\tau_{i,j}(k)$ and $\alpha_{i,j}(k)$ as formulated by Equation 2.

$$\phi_{i,j}(k) = \tau_{i,j}(k) - \alpha_{i,j}(k) \tag{2}$$

De facto red exists when the signal is green but traffic cannot proceed because of backed-up traffic on a receiving street, as indicated in Figure 2*b*. To avoid the de facto red situation, the effective green

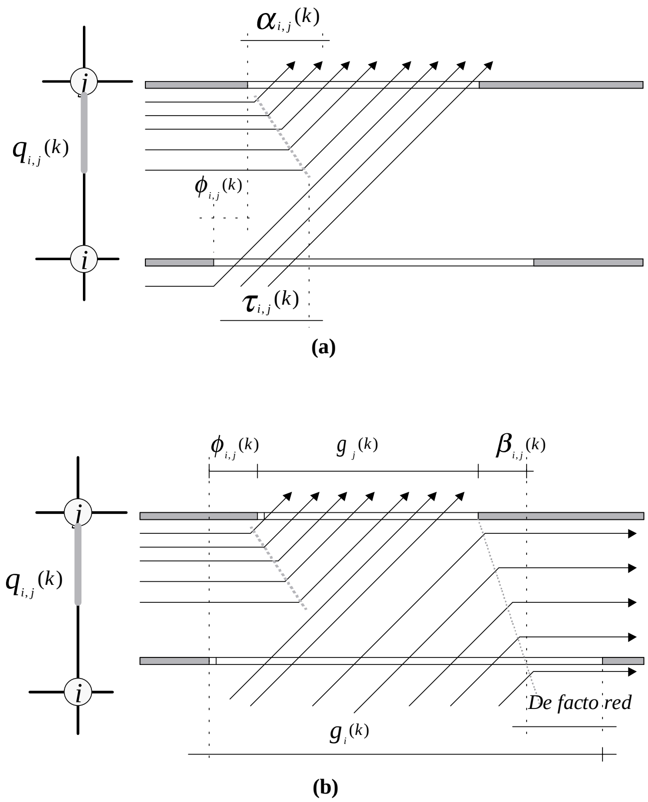


FIGURE 2 (*a*) Ideal offsets between two adjacent signals. (*b*) De facto red at upstream signals.

time for the upstream signal $g_i(k)$ should be less than the sum of the effective green time for the downstream signal $g_j(k)$, the offset between the two signals $\phi_{i,j}(k)$, and the time it takes for a stopping shock wave to propagate upstream, $\beta_{i,j}(k)$. This mechanism is formulated by Equation 3.

$$g_i(k) \le g_j(k) + \phi_{i,j}(k) + \beta_{i,j}(k) \tag{3}$$

Green Time Constraints: Crossing Traffic

Equation 4 provides a formulation to calculate effective green times at a downstream signal serving movements that cross coordinated arterials $g_j^c(k)$.

$$g_j^c(k) = \text{ext}\ \phi_{i,j}(k) + \phi_{i,j}(k+1) - \left[g_j(k) + \Delta\right] - \Delta \qquad (i, j) \in L_p \tag{4}$$

This equation requires knowledge of the effective green time that serves through (coordinated) movements $g_j(k)$, the ideal offset $\phi_{i,j}(k)$, the extended offset ext $\phi_{i,j}(k)$ between two adjacent signals, and lost time Δ.

The extended offset is a measure of neighboring intersection linkage over time, and it reveals the relationship between a cycle length at an upstream intersection, $C_i(k)$, and the ideal offset at the next cycle $\phi_{i,j}(k+1)$. The extended offset is formulated by Equation 5.

$$\text{ext}\ \phi_{i,j}(k) = C_i(k) + \phi_{i,j}(k+1) \qquad (i, j) \in L_p \tag{5}$$

Lock-In Loop Constraints

Signal coordination has to comply with a set of constraints that are determined by the structure of a signal network. The sum of offsets and green times around any loop of the network is equal to a multiple of the cycle time (*5*). This is called a lock-in loop constraint. Consider a directed loop in Figure 1*a* that traverses counterclockwise from Signal 10 to 5 to 4 to 9, and to 10. Figure 1*b* presents a typical time–space diagram, and Equation 6 indicates the lock-in constraint for this loop. The superscripts indicate phase number; thus, for Signal 10, the green time for Phase 1 (east–west) is indicated by $g_{10}^1(k)$ and that for Phase 2 (north–south) is indicated by $g_{10}^2(k)$.

$$\begin{aligned}&\phi_{10,5}(k) - \phi_{5,4}(k) + \phi_{4,9}(k) - \phi_{9,10}(k) + \left[g_5^2(k) + \Delta\right]\\&+\left[g_4^1(k) + \Delta\right] + \left[g_9^2(k) + \Delta\right] + \left[g_{10}^1(k) + \Delta\right] = \left[g_{10}^1(k) + \Delta\right]\\&\qquad + \left[g_{10}^2(k) + \Delta\right] + \left[g_{10}^1(k+1) + \Delta\right] + \left[g_{10}^2(k+1) + \Delta\right]\end{aligned} \tag{6}$$

For the network presented in Figure 1*a*, there are 12 loops and 20 intersections; thus, there are 2×20 lock-in constraints. In general, the lock-in constraint is formulated by Equation 7. $N(l)$ is a set of nodes of directed loop l. $F(l)$ is the set of forward links in loop l, where traffic flow moves in the same direction as the loop's direction. $R(l)$ is the set of reverse links in loop l, where traffic flow moves in the direction opposite the loop's course. $g_{j,l}(k)$ is the green time of signal j at cycle k serving movements in loop l. n is an integer number and P is the number of loops in the network.

$$\sum_{i,j \in F(l)} \phi_{i,j}(k) - \sum_{i,j \in R(l)} \phi_{i,j}(k) + \sum_{j \in N(l)} g_{j,l}(k) = \sum_{k,j \in N(l)}^{k+n} C_j(k) \qquad \forall l \in P \tag{7}$$

The ideal offset is assigned only between two signals along coordinated arterials. For BSNP, not all offsets in Equation 7 are ideal. For the loop formed by Signals 10, 5, 4, 9, and 10 (see Figure 1), all movements are controlled by ideal offsets, thus satisfying Equations 2 and 7, except for a movement from Signal 4 to 9, which uses the implied offset satisfying only Equation 7.

Storage Capacity Constraints: Noncoordinated Arterials

The queue along noncoordinated arterials must be less than the storage capacity of approach links. This constraint is to ensure that blockages on crossing (coordinated) arterials do not occur.

$$q_{i,j}(k) \le \max q_{i,j} \qquad (i, j) \in L_s, k = 1, \ldots, K \tag{8}$$

Queue limitation along secondary movements is formulated by Equation 8, where max $q_{i,j}(k)$ is the maximum number of vehicles in the queue that can be stored in the approaching links between signals i and j. L_s is a set of streets carrying noncoordinated movements.

State Equations

During oversaturation, the number of vehicles in a queue at the beginning of a cycle depends on the queue in the preceding cycle. To account for this interdependency, queue formation and dissipation at a different approach in the system are tracked for every time step taken or for every cycle. The queue at intersection j at the beginning of the cycle $k + 1$, $q_{i,j}(k + 1)$, is calculated with Equation 9, where $A_{i,j}(k)$ is the number of vehicles arriving during that cycle.

$$q_{i,j}(k + 1) = q_{i,j}(k) + A_{i,j}(k) - D_{i,j}(k) \qquad \forall j \in N, (i, j) \in L \tag{9}$$

Boundary Values

All control parameters should be within reasonable ranges. Equation 10 formulates the allowable green time where $g_{\min}$ is the lower bound and $g_{\max}$ is the upper bound, and Equation 11 defines the initial values for queues at $k = 0$.

$$g_{\min} \le g_j(k) \le g_{\max} \qquad \forall j \in N; k = 1, \ldots, K \tag{10}$$

$$q_{i,j}(k) = q_0 \qquad k = 0; \forall(i, j) \in L; k = 1, \ldots, K \tag{11}$$

Departure Volumes: Coordinated Movements

For a given cycle, traffic volume that departs from signal j depends on the effective green time available serving coordinated movements $g_j(k)$, the saturation flow rate s_j, queue $q_{i,j}(k)$, and arrival flow $A_{i,j}(k)$. Depending on whether all green is used at the respective upstream intersections $g_i(k)$, the discharge at signal j is determined by Equation 12.

$$D_{i,j}(k) = \min\left\{\begin{array}{l} q_{i,j}(k) + A_{i,j}(k) \\ q_{i,j}(k) + \left(\dfrac{A_{i,j}(k)}{g_i(k)}\right) \times \left(g_j(k) - \dfrac{q_{i,j}(k)}{s_j}\right) \end{array}\right\} \qquad \forall j \in N \tag{12}$$

This equation dictates that the number of vehicles discharged during green time depends on the storage that is available on the approaches of downstream intersections. If the arrival exceeds the available storage capacity, the traffic departed from intersections is limited by the available capacity. As a result, queues will accumulate. If the capacity is larger than the demand, both the arrivals and the existing queue will be processed.

Departure Volume: Noncoordinated Movements

Two conditions can occur when nonideal offsets are applied: wasted green and blockages. The duration of wasted green is a period of unused green time after the last vehicle in a queue is cleared until new arrivals pass intersections. The wasted green can occur if the start green is set too early relative to the upstream green initiation. In other words, green time is wasted when the magnitude of implied offsets is larger than the ideal offsets. Unused green time can also occur when excessive green time is assigned. Blockages, on the other hand, can occur if traffic is prevented from leaving upstream intersections because of the presence of a queue at downstream intersections. Blockages form when the stopping shock wave, caused by the midblock traffic, reaches upstream intersections and ends when the blockage is cleared. During this time, no vehicles are processed. A combination of the occurrence of wasted green time and blockages determines the departure flows along noncoordinated traffic movements.

GA-BASED PROCEDURES TO SOLVE BSNP

Consider a set of genetic solutions represented as S. It consists of all green times serving coordinated movements—that is, the northbound ($p_{20,5}$) and all arterials running east–west. This set of solutions is formulated in Equation 13, where $g_i^1(k)$ is green time for Phase 1 at signal i, and $g_i^2(k)$ is green time for Phase 2; it is indicated schematically by solid arrows in Figure 1*a*. At Signal 20, green times allocated for all coordinated movements are included in the solution. This is necessary to assign a reference signal at which the cycle length is known. Note that not all green times are included in the genetic solution. As indicated in Equation 13, decision variables included in S for one cycle are $m = 21$. If the duration of oversaturation is $K = 15$ cycles, then there are $mK = 21 \times 15 = 315$ decision variables for all cycles. These green times determine the number of vehicles processed by the network, accumulated queues, offsets, and all remaining green times.

$$\boldsymbol{S} = \left\{\begin{array}{l} g_{20}^1(k), g_{20}^2(k), g_{15}^2(k), g_{10}^2(k), g_5^2(k), \\ g_1^1(k),\ g_2^1(k),\ g_3^1(k),\ g_4^1(k), \\ g_6^1(k),\ g_7^1(k),\ g_8^1(k),\ g_9^1(k), \\ g_{11}^1(k),\ g_{12}^1(k), g_{13}^1(k), g_{14}^1(k), \\ g_{16}^1(k),\ g_{17}^1(k), g_{18}^1(k), g_{19}^1(k), \end{array}\right\} \qquad k = 1, \ldots, K \tag{13}$$

The signal coordination problem is solved first by using the genetic initial solution randomly generated. Queues for all signal approaches at the beginning of the first cycle's green time are given. For each cycle, queues, offsets, and the number of vehicles processed by signals along the coordinated northbound arterial $p_{20,5}$ are calculated with Equations 2, 9, and 12. This coordinated northbound arterial is called the closing path because it locks in all other east–west offsets. Given these queues and offsets, the green times of the closing path that serve the crossing arterials are then calculated with Equations 3, 4, and 5. Next, queues, offsets, and the number of vehicles processed by all coordinated arterials running east–west are determined. Finally, all operational parameters for noncoordinated arterials are calculated. The procedure for implementing GA is summarized as follows:

1. Define genetic solutions S as formulated in Equation 13.
2. Let $k = 0$ (cycle 0).
3. Calculate departure flows at signals along the closing path.
4. Calculate crossing green times at signals along the closing path.
5. For all coordinated arterials that cross the closing path, calculate
   –Departure flows at signals along the arterials, and
   –Crossing green times except for the intersections between the arterials and the closing path, as they have been calculated in the previous step.
6. Calculate all departure flows along noncoordinated arterials.
7. If $k = K$, proceed to Step 8; otherwise $k = k + 1$, and proceed to Step 3.
8. Calculate the total number of vehicles exiting from each signal.
9. Calculate the GA's fitness function, as formulated in Equation 1.

Given the initial green times, the calculation of departure flows at signals is straightforward. It depends primarily on the arrival rate, the queues at the beginning of the cycle, and the duration of the available green time at signals. Calculation of crossing green times, on the other hand, is somewhat complicated. Consider, for example, a two-signal system along the closing path: Signals 20 and 15.

Defined by the genetic solutions, the green times of the two phases at Signal 20 are known; that is, $g^1_{20}(k)$ and $g^2_{20}(k)$ (see Figure 1*a*). Signal 20 is a reference signal with a known cycle length. The green time at Signal 15 serving northbound is also known, $g^2_{15}(k)$. Offsets at cycles k and $k + 1$ between these two signals can be calculated because we can calculate queues at the end of cycles k and $k + 1$, provided that the departure rates from the two signals are known (calculated with Equation 12). These offsets are ideal, thus providing a progression for the northbound movement. With Equations 4 and 5, the first phase of green time at Signal 15, $g^1_{15}(k)$, is calculated as follows:

$$\begin{aligned} g^1_{15}(k) &= \text{ext}\phi_{20,15}(k) + \phi_{20,15}(k+1) - \left[g^2_{15}(k) + \Delta\right] - \Delta \\ &= \left[C_{20}(k) - \phi_{20,15}(k)\right] + \phi_{20,15}(k+1) - g^2_{15}(k) - 2\Delta \end{aligned} \tag{14}$$

This process of calculation is repeated until all crossing green times along the closing path are calculated—namely, $g^1_{10}(k)$ and $g^1_5(k)$. Having calculated all signal timings for signal coordination on this path, one can design signal coordination on arterials running east–west. Signals along these arterials that belong to the closing path function as reference signals with known cycle lengths. Signal 15 becomes a reference signal for coordinated path $p_{11,15}$. Similarly, Signal 10 is a reference for $p_{10,6}$, and Signal 5 is a reference for $p_{1,5}$. Starting from the reference signals, a similar analysis can be made for these arterials, as calculated for the closing path. Note that Signal 20 is a reference signal for two different coordinated paths: a northbound arterial ($p_{20,5}$) and an eastbound arterial ($p_{20,16}$). Indeed, this signal acts as a reference signal for all signals in the network.

Until now, the magnitude of signal timings, including green times and offsets, has been defined, except for the offsets along noncoordinated (north–south) arterials. These offsets are not ideal, and traffic progressions are not expected along these arterials. Accordingly, offsets can be defined by the geometry of signal networks satisfying Equation 7. For an illustration of this process, consider a loop of Signals 10, 5, 4, and 9 in Figure 1*a* and its time–space diagram as indicated in Figure 1*b*. Equation 6 is used to calculate the offsets for noncoordinated arterials. All offsets in the lock-in constraint, except for the offset between Signals 4 and 9, have been set ideal, hence promoting traffic progression. At this time, all the green times indicated in the preceding equation are known. The offset between Signals 4 and 9 is calculated as follows:

$$\begin{aligned} \phi_{4,9}(k) &= \sum_k^{k+1} C_{10}(k) - \begin{pmatrix} \phi_{10,5}(k) - \phi_{4,5}(k) - \phi_{10,9}(k) + g^2_5(k) \\ + g^1_4(k) + g^2_9(k) + g^1_{10}(k) - 4\Delta \end{pmatrix} \\ &= \left(\sum_k^{k+1} C_{10}(k) - g^1_{10}(k) + \phi_{10,9}(k) - g^2_9(k) - 2\Delta\right) \\ &\quad - \left[\phi_{10,5}(k) + g^2_5(k) - \phi_{4,5}(k) + g^1_4(k) + 2\Delta\right] \end{aligned} \tag{15}$$

Note that the first term on the right-hand side of the equation is the switching time of Signal 9 for the green duration at cycle k, and the second term is that of Signal 4. This relationship can be observed clearly from Figure 1*b*. The offset between any pair of signals, for any cycle, along noncoordinated arterials can be calculated similarly.

The GA's fitness values are calculated equal to the objective functions, as formulated in Equation 1. A number of procedures are used to incorporate equality and inequality constraints into the GA search. The equality constraints (Equations 2, 4, 5, 7, 9, 11, and 12) are included in the model of the algorithm. The inequality constraints (Equations 3, 8, and 10) are dealt with through the penalty method; that is, the rank of GA fitness function is degraded in relation to the degree of constraint violations. The constrained problem described in the preceding section is transformed to an unconstrained problem by associating penalty with all inequality constraint violations.

APPLICATION OF μGA PROCEDURES

The algorithm is tested on a one-way arterial system with 20 intersections as indicated in Figure 1*a*. Two arterials are assumed to carry very heavy traffic: 2,000 vehicles per hour per lane (vphpl). These are the northbound arterial from Signal 20 to Signal 5 and the eastbound arterial from Signal 10 to Signal 6. Furthermore, eastbound traffic enters the network at Signal 1 and Signal 11 with a flow rate of 1,800 vphpl, whereas westbound traffic enters the network at Signal 20 with a flow rate of 1,800 vphpl. These are defined as major movements. In addition, all minor traffic enters the network with a flow rate of 1,500 vphpl at Signals 2 and 4 for the southbound traffic and at Signals 16 and 19 for northbound traffic. These entry flows are assumed constant during an oversaturated period. For coordinated movements, $g_{\max} = 90$ and $g_{\min} = 30$ s; for noncoordinated movements, $g_{\max} = 60$ and $g_{\min} = 20$ s. The number of arterial lanes is two. Speed limit, or desired speed, equals 40 ft/s, and vehicle acceleration is 4 ft/s^2. Initial queues at all coordinated approaches are 20 vehicles per lane. Effective vehicle length is 25 ft. Starting and stopping shock-wave speeds are 16 and 14 ft/s, respectively.

Instead of working with a single solution, GA works with a number of genetic solutions or individuals collectively known as a population. The individual consists of a set of green time variables to be optimized, as formulated in Equation 13. Every variable is coded in a fixed-length of binary substring d, allowing each variable to have 2^d possible values. The length of a substring is decided by the precision needed for the green time. Five seconds is chosen to be the smallest unit of green time. Therefore, the total number of possible green time values must be $(g_{\max} - g_{\min})/5 = (90 - 30)/5 = 12$. This number can be set equal to 2^d, and d can be computed as $\log_2(12)$; d should equal at least four bits of binary string. Each green time is represented as a four-bit binary code. For $mK = 21 \times 15 = 315$ variables, the total string length is $315 \times 4 = 1{,}260$ for each individual. A substring with all zeros represents the minimum green time $g_{\min}$ and that with all ones represents the maximum green time $g_{\max}$. Any other substring is decoded to a green time, g, as formulated by Equation 16, where DV is the decoded value of a substring. Thus, the decoded value for substring 1,001 is $(1 \times 2^3) + (0 \times 2^2) + (0 \times 2^1) + (1 \times 2^0) = 9$, and the corresponding green time for such a substring is $30 + [(90 - 30)/(2^4 - 1)] \times 9 = 66$ s.

$$g = g_{\min} + \frac{g_{\max} - g_{\min}}{2^d - 1} DV \tag{16}$$

The number of individuals in a population determines the quality of solutions and the duration of GA computation time. The optimal population size equals the square root of the size of individuals, $n = \sqrt{1{,}260} = 36$ (*6, 7*). In this GA procedure, the tournament selection scheme is applied on a population in which two or more individuals are chosen at random for a tournament, and according to the individual's fitness value, the better individual is selected and obtains two copies in the mating pool (*8*). Crossover is applied to the individual of the mating pool. Two individuals are picked from the mating pool at random, and they are cut at a randomly selected site(s).

The portions of both individuals bracketed by this site(s) are swapped among themselves to create two new individuals. The basic procedure of μGA is as follows (*9*):

1. Select a population of size p randomly, or $p - 1$ randomly, and one good individual from any previous search.
2. Evaluate fitness values and determine the best individual, and carry it to the population of the next generation (elitist strategy).
3. Determine the remaining individuals by using genetic operators (tournament selection and crossover).
4. Check for convergence—that is, when the highest fitness value is sufficiently close to the average fitness of the population. If the search converges, proceed to Step 1 (one epoch); otherwise, proceed to Step 2.

The process of genetic operation continues until the prescribed number of epochs is reached. The μGA was executed for 100 epochs, resulting in about 5,000 generations. The completion of a GA search required about 50 min on a PC 333-MHz machine.

DISCUSSION OF RESULTS

Cycle Lengths

Figure 3*a* and 3*b* present the fluctuation of cycle lengths of signals along two coordinated arterials intersected at Signal 15 (see Figure 1*a*). Figure 3*a* indicates cycle lengths for the northbound arterial $p_{20,5}$, or the closing path; Figure 3*b* indicates the cycle length for the eastbound arterial $p_{11,15}$. Signal 15 is the farthest downstream and the critical signal for the eastbound arterial. The thickest line with empty circles in both figures indicates the fluctuation of cycle time for Signal 15. A larger difference of cycle lengths during the first few cycles indicates the period in which signals reduce and dissipate queues. Once queues are cleared or reduced, the algorithm starts setting the signal timing in synchrony with a common cycle length. Note that a common cycle for the closing path is achieved at earlier cycles, whereas that for the eastbound arterial occurs after Cycle 6.

Unlike the eastbound arterials, the westbound arterials intersect the closing path at the farthest upstream signals. Figure 3*c* presents the fluctuation of the cycle length along the westbound arterial $p_{20,16}$. This arterial intersects with the closing path at Signal 20, which serves as the critical signal for the arterial. Comparing that with Figure 3*b* verifies that a set of common cycles is achieved at an earlier stage for the westbound arterials. This is expected because the westbound arterial is relatively less restrictive because the critical signal (Signal 20) is located at the entry point. At this signal, the algorithm sets a lower cycle time at the first few cycles to restrain entering traffic, thus promoting an earlier queue dissipation process at signals.

Green Time Allocation

Close examination of the changes of green times demonstrates the capability of the algorithm in managing queues and providing desirable traffic progression. All signals along the east–west arterials are coordinated for single traffic progression, except for the critical (exit or entry) signals whose signal timings are set to provide traffic progression for two competing movements (see Figure 1*a*). Figure 4 presents the dynamics of signal timings along the westbound arterial $p_{10,6}$. Signal 10 is the critical signal for this arterial. Note that, for a given cycle, the green time serving the westbound progression (g_1) on average increases as traffic progression occurs in the downstream direction. This is required for the arterial to dissipate queues in achieving normal progression. Queue dissipation occurs from the farthest downstream signals and propagates upstream. Upstream signals control the queues for downstream signals to process them and thus avoid spillback.

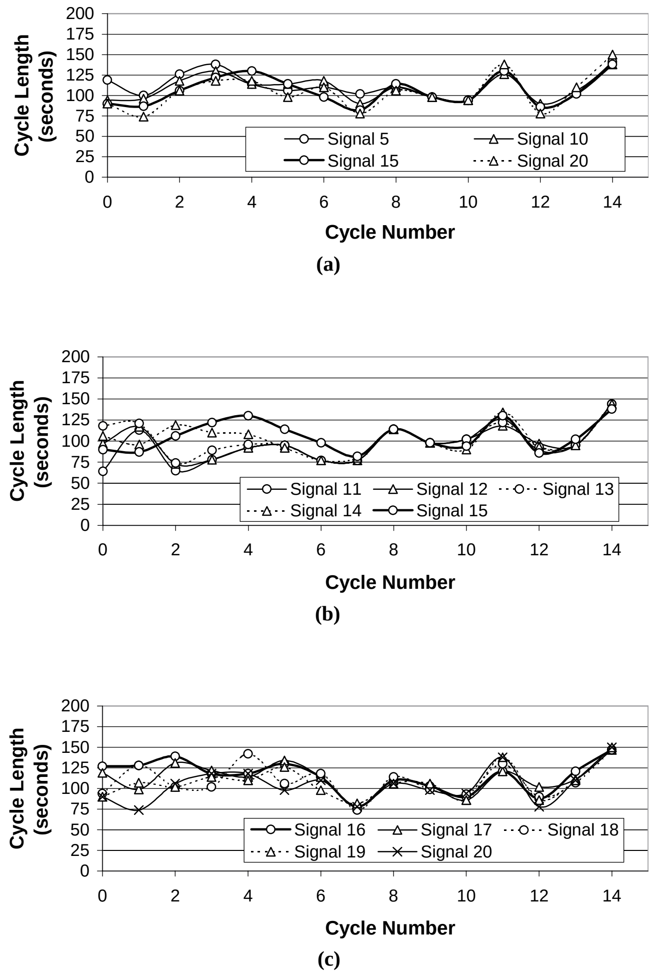


FIGURE 3 Fluctuation of cycle lengths along coordinated arterials: (*a*) northbound, or the closing path; (*b*) eastbound; (*c*) westbound.

Figure 5 presents the dynamics of signal timings on the arterial with eastbound progression $p_{11,15}$. Unlike the westbound arterials with critical signals at the entry point, the exit signal (Signal 15) for the eastbound arterial $p_{11,15}$ is the critical one; thus, more restrictive green times (g_1) on the upstream signals are set during the first few cycles. As a result, the upstream signal system releases fewer vehicles in the platoon. The algorithm maintains the released platoons at a manageable level so that queues at this busiest signal do not grow excessively. Once queues have been cleared up or reduced at all signals, green times (g_1) along the arterial increase and fluctuate following the green time pattern of the critical signal.

Queue and Offsets

Figure 6*a* indicates how queues and offsets are interrelated and how the load of traffic is distributed among signals along the eastbound

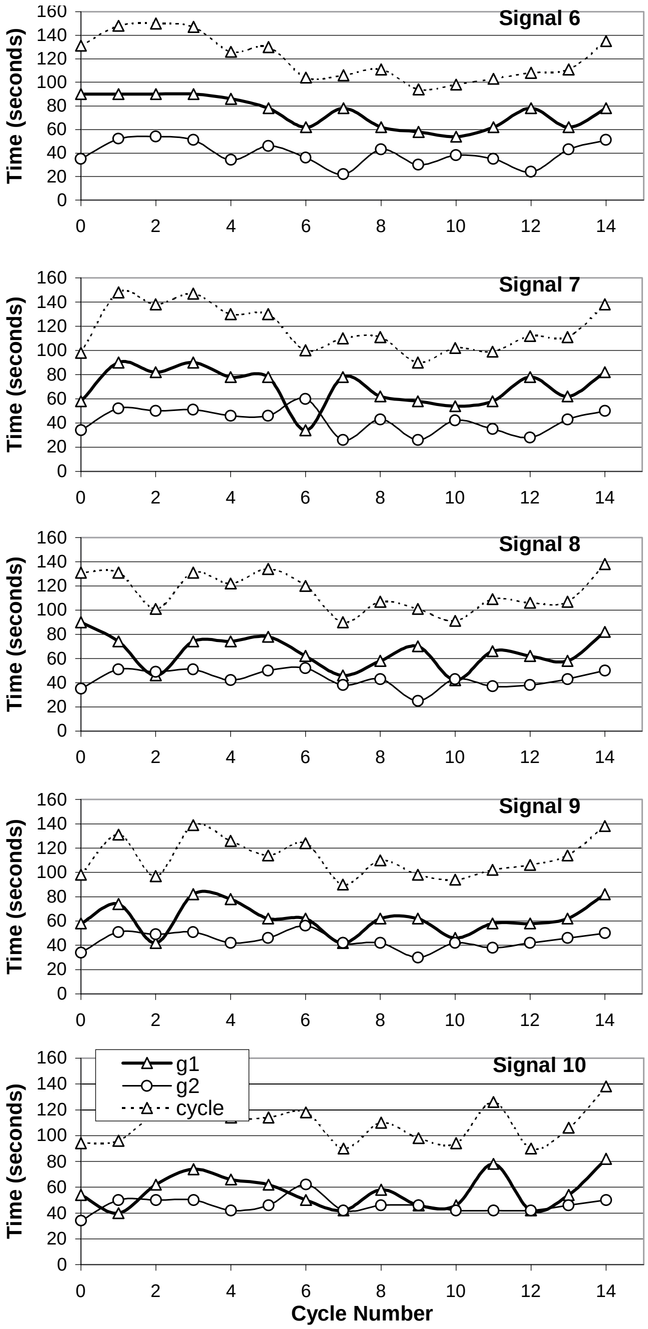


FIGURE 4 Fluctuation of green times and cycle lengths for westbound coordinated arterial from Signal 10 (the critical signal) to Signal 6: westbound traffic progression is controlled by westbound green time (g_1).

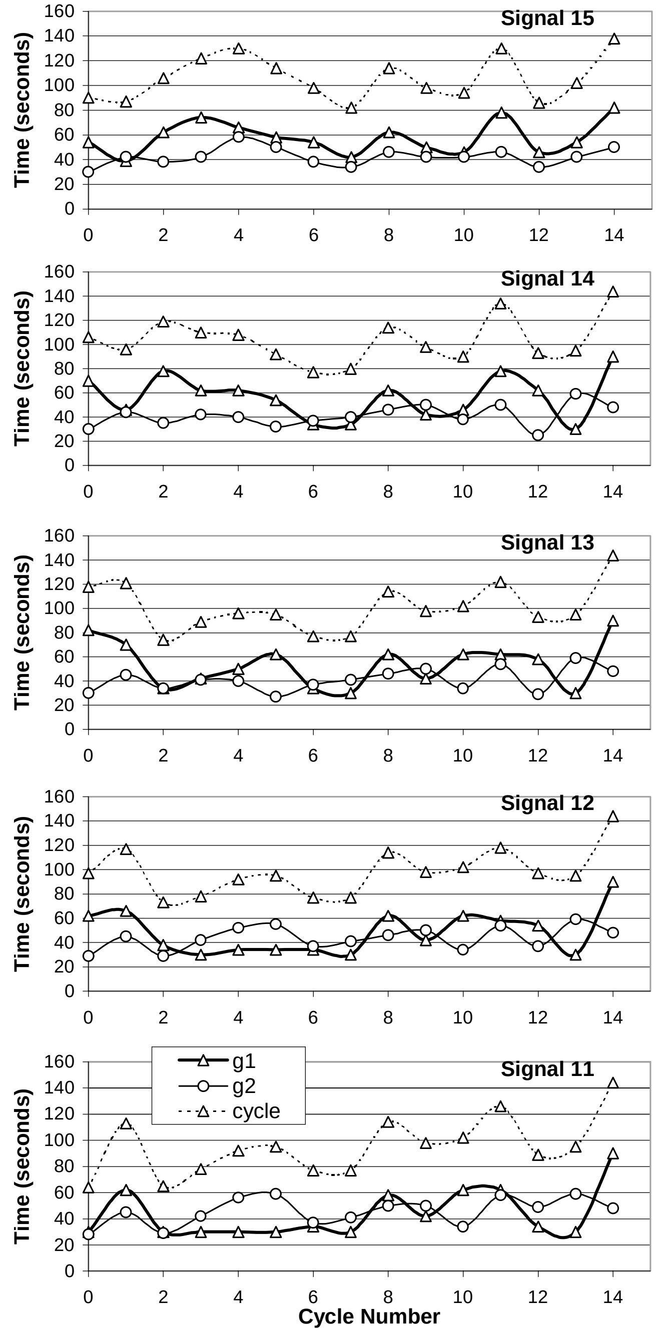
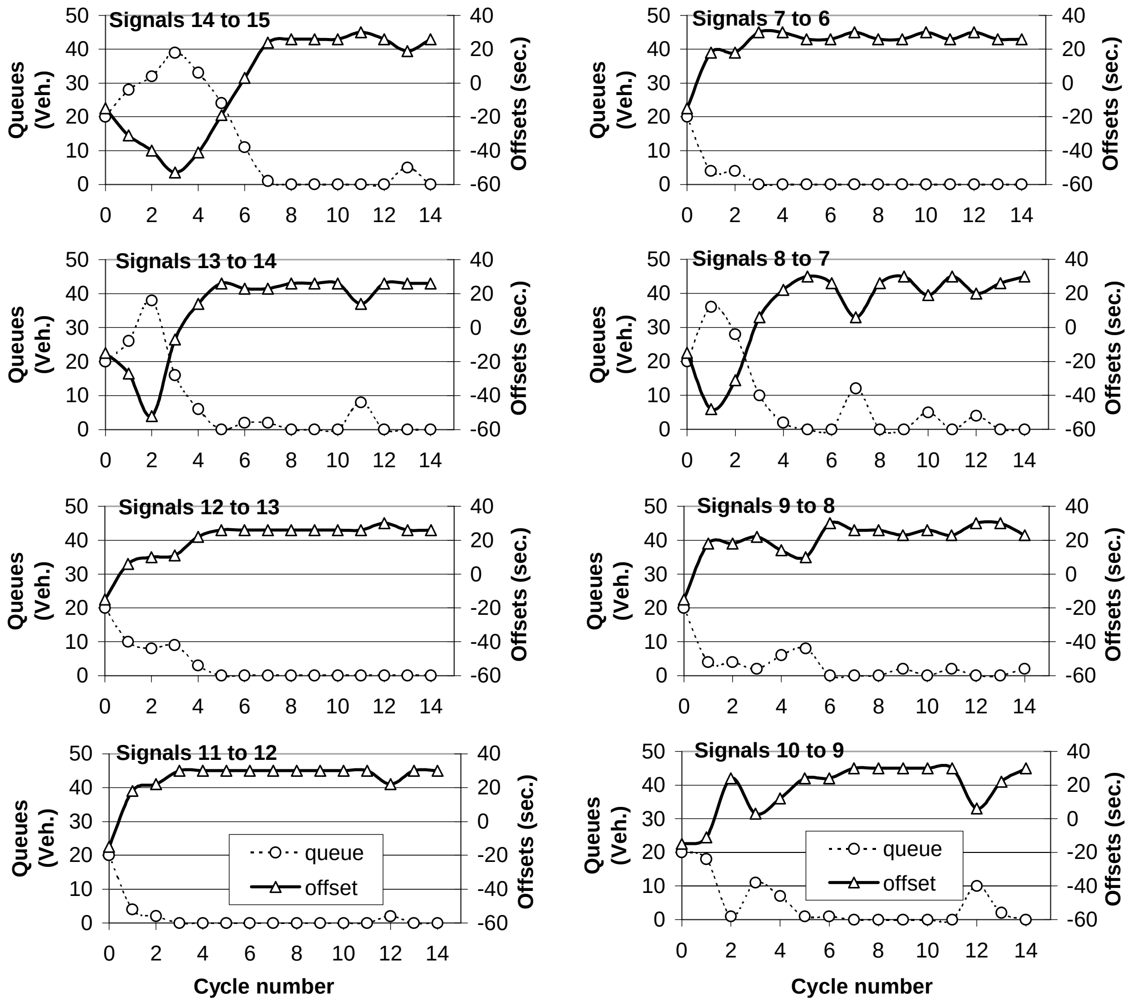


FIGURE 5 Fluctuation of green times and cycle lengths for eastbound coordinated arterial from Signal 11 to Signal 15 (the critical signal): eastbound traffic progression is controlled by eastbound green time (g_1).

arterial $p_{11,15}$. The heavy load of the queue-dissipation process at Signal 15 (critical signal) postpones the green initiation of Signal 14 (upstream signal) by setting a negative offset, thus causing an accumulated local queue along the approach to Signal 14. This deferral takes place before Cycle 6, when local queues at the approach to Signal 15 are being cleared up. The mechanism of load shifting among signals propagates upstream. Note that the positive offsets are set at an earlier stage of cycles in the upstream direction. For example, an offset of 20 s is set after Cycle 7 at Signal 15, after Cycle 5 at Signal 14, after Cycle 4 at Signal 13, and after Cycle 3 at Signal 12. However, this does not mean that the queue-dissipation process propagates to downstream signals. In fact, it is the other way around. The reason why early positive offsets are set at upstream signals is that the algorithm restrains entering traffic by allocating close to the minimum green times at the entry signals, which in turn releases existing local queues downstream.

In contrast, the queue-offset relation along the westbound arterial $p_{10,6}$ shows that earlier positive offsets (forward traffic progres-



(a) (b)

FIGURE 6 Dynamics of offsets and queues: (*a*) eastbound coordinated arterial with Signal 15 as the critical signal; (*b*) westbound coordinated arterial with Signal 10 as the critical signal.

sion) and the queue-dissipation process occur at the downstream signals, as indicated in Figure 6*b*. This is expected because the exit signal of this coordinated arterial is not the critical one and is not as restricted as the exit signal on eastbound arterials. Therefore, the algorithm can release more vehicles at the entry signal (Signal 10), and any queues at downstream signals can be released faster than the time it takes for vehicles (discharged at the entry signals) to travel to the exit signals. Consequently, the positive offsets are achieved earlier at downstream signals.

In general, forward traffic progression is achieved after several cycles of a signal-coordination process. Depending on the magnitude of the local queues, the algorithm sets the appropriate offsets to dissipate local queues during the early cycles and sets less green time to control arrivals. The magnitude of local queues and the position of critical signals determines the beginning of forward progression. When the critical signal is located at the exit point, the progression is expected to occur at a later stage; when it is located at the entry point, the progression tends to occur at an earlier stage. This mechanism of traffic progression is expected because a system with a critical signal at the exit point operates in a more restrictive environment. The green bandwidth along coordinated arterials also varies over the cycles, depending on the existence of local queues and on the magnitude of arrival volumes. Because of space limitations, green bandwidth is not discussed here. This variation of green bandwidth reflects the ability of the algorithm to effectively allocate green times. An inadequate or wasted green time at any signal would jeopardize the performance of the whole signal network during an oversaturated period. Green time must be used productively and therefore is reflected in the variation of the green bandwidth.

CONCLUSION AND RECOMMENDATIONS

This paper presents an algorithm to design signal coordination for networks with oversaturated intersections. The algorithm successfully provides signal timing that is responsive to the temporal variation of arrival volumes and queues at the intersection approaches. It generates an online load-balancing mechanism in which critical intersections are protected from becoming oversaturated. When critical intersections were located at exit points of the network, during the first few cycles, all upstream signals managed volumes of entering traffic by setting lower green times. This was combined with setting negative offsets of longer duration at the exit signals. Later, the positive offsets were gradually set as the algorithm promoted forward green bands. When critical signals were located at entry points, the negative offsets were maintained for longer duration at the entry signals. This was done to ensure that all local queues were cleared before more vehicles arrived at downstream signals. The algorithm also provided a set of common cycles that are necessary for signal coordination. The common cycle changed depending on arrival volumes and the optimal effective green times for the queue-dissipation process.

This algorithm distributed queues temporally and spatially when signal networks were oversaturated. The algorithm explicitly recognized physical limitations of the approaches for storing vehicles and prevented blockage of upstream intersections. The signals shared the burden of excessive traffic and simultaneously provided smooth traffic movement along user-defined corridors. Further studies are needed to expand the algorithm for signal networks that involve loops of coordinated arterials. In addition, the algorithm should accommodate the demand for turning movements, different types of queue-management strategy, and two-way arterial networks. To reduce GA computation time, μGA needs to be modified so it can be executed in parallel computing machines. Furthermore, analysis of the scalability of the algorithms for large networks is needed.

REFERENCES

1. Abu-Lebdeh, G. *Development of Dynamic Traffic Signal Control Procedures for Oversaturated Arterials and Genetic Algorithms Solutions.* Ph.D. dissertation. University of Illinois at Urbana–Champaign, 1999.
2. Abu-Lebdeh, G., and R. F. Benekohal. Development of Traffic Control and Queue Management Procedures for Oversaturated Arterials. In *Transportation Research Record 1603,* TRB, National Research Council, Washington, D.C., 1997, pp. 119–127.
3. Lieberman, E., J. Chang, and E. S. Prassas. Formulation of a Real-Time Control Policy for Oversaturated Arterials. Presented at 79th Annual Meeting of the Transportation Research Board, Washington, D.C., 2000.
4. Goldberg, D. E. *Genetic Algorithm in Search, Optimization, and Machine Learning.* Addison Wesley Longman, Reading, Mass., 1989.
5. Gartner, N. H. Constraining Relations Among Offsets in Synchronized Signal Networks. *Transportation Science,* Vol. 6, 1972, pp. 88–93.
6. Abu-Lebdeh, G., and R. F. Benekohal. Convergence Variability and Population Sizing in Micro-Genetic Algorithms. *Computer-Aided Civil and Infrastructure Engineering,* Vol. 14, 1999, pp. 321–334.
7. Harik, G., E. Cantu-Paz, D. E. Goldberg, and B. Miller. The Gambler Ruin Problem, Genetic Algorithm, and the Sizing of Populations. *Evolutionary Computation,* Vol. 7, No. 3, 1999, pp. 7–12.
8. Goldberg, D. E., and K. A. Deb. A Comparative Analysis of Selection Schemes Used in Genetic Algorithms. In *Foundations of Genetic Algorithms* (G. J. E. Rawlins, ed.), Morgan Kaufmann, San Mateo, Calif., 1991, pp. 69–93.
9. Krishnakumar, K. Micro-Genetic Algorithms for Stationary and Nonstationary Function Optimization. In *Proc., SPIE, Vol. 1196, Intelligent Control and Adaptive Systems,* Philadelphia, Pa., 1989, pp. 289–296.

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