THE GENERALIZED COMBINATION METHOD
FOR AREA TRAFFIC CONTROL

by

Nathan Gartner and John D.C. Little

OR 020-73 August 1973

Operations Research Center
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Research supported by KLD Associates, Inc., in connection with Department of Transportation Contract FL-11-7924.
ABSTRACT

A procedure is described for determining optimal signal settings in a network including cycle time, splits of green time and offsets. A number of cycle times are scanned in the process. For each cycle, splits are determined locally at each intersection according to proportions of conflicting traffic loads, and offsets are optimized by the Generalized Combination Method. Total cost for travelling through the signal-controlled intersections in the network is evaluated as the sum of two components: deterministic and stochastic. The deterministic component is a function of the offsets throughout the network and generally increases with cycle length. The stochastic component is dependent on the expected overflow queue on each link and decreases with cycle length. Optimal settings are determined as an equilibrium point of minimum total costs resulting from the combined effect of the two components.
INTRODUCTION

The primary objectives of an area-wide traffic control system are to provide smooth flow conditions for all traffic streams through the area and to reduce the delay, or travel time, incurred by the users of the system. The variables of each signal program that affect the traffic flow are: cycle time, splits of green time and offsets. A coordinated traffic signal network requires a common cycle time for all signals in the network, or a cycle which is a submultiple of some master cycle. In some cases it is advantageous to partition the network into subnetworks that may operate with nonsynchronous cycle times. The conventional procedure for determining the control variables is to use a sequential decision process: A common cycle time is selected for the network first. Then the splits at each intersection are determined according to the proportions of demand/capacity ratios on conflicting approaches. Finally, linking of the signals is achieved by an appropriate method for selection of a fundamental set of offsets throughout the network.

Experience of researchers and practitioners in the urban traffic control field has shown that cycle time may well be the one most important control variable in a synchronized traffic signal network (1). The approaches for selecting a cycle time can be divided into two classes. The first class is the node approach. Since through capacity increases with cycle length, this approach is based on analysing the capacity requirements of each intersection in the network. The common cycle time is determined according to the requirements of the most heavily loaded intersection, i.e., the intersection with
the highest sum of demand/capacity ratios on conflicting signal phases. A
procedure that is used for a single intersection, such as Webster's
method(2), is then used to calculate the cycle length. This approach has
been primarily used in conjunction with offset optimization methods such as
COMBINATION and TRANSYT(3). The main deficiency in this approach is that the
interaction of flows in the spatial road network structure of the area is
disregarded. A formula devised for an isolated intersection,
assuming randomly distributed arrival times of cars, is not necessarily valid
in a network situation where flows are fed from adjacent intersections. The
result is generally a cycle time that is too long, thus causing excessive
delays (see also reference 4).

The second class is the network approach. In this case an attempt
is made to select a cycle time that, while satisfying the capacity require-
ments at each intersection, is also congruent with the particular network
structure at hand. Simple examples in this category are the arterial pro-
gression schemes in which a cycle that produces maximal bandwidths is selected
according to distance and speed data (e.g., 5,6,7). The underlying principle
is that the optimal progression (i.e., offsets between signals), for a given
block-length pattern, is strongly dependent on cycle time. In a general net-
work this approach is principally used by SIGOP(8,9). A predetermined number
of cycle times are scanned in this method. For each cycle, offsets are
optimized by the OPTMIZ subroutine and performance is evaluated by a coarse
simulation of traffic flow through the network (the VALUAT subroutine). The
optimal set of cycle and offsets is selected according to the results obtained
by VALUAT. TRANSYT indicates also the possibility to iterate on cycle time
in conjunction with its hill-climbing procedure for offset selection(10).
However, computational considerations seem to rule out this possibility in practice. Two deficiencies of the network approach in SIGOP are apparent: first, the offset optimization procedure determines a local optimum rather than a global optimum and, second, stochastic effects on link performance are ignored. These effects do not affect the selection of offsets at a fixed cycle time, but are of prime importance in evaluating a range of cycle times. They become pronounced as a signalized intersection approaches its capacity and in an optimal procedure would deter the cycle time from assuming values close to the minimum. One typical study(11) has shown that the lower bound on cycle time was consistently selected as the optimal value. Stochastic effects would have conceivably shifted the result upwards.

In this paper, network settings (including cycle, splits and offsets) are determined in conjunction with a rigorous synchronization procedure (i.e., a procedure capable of determining the global optimum), the Generalized Combination Method (GCM). The Combination Method (CM), which is an offset optimization procedure applicable to series-parallel networks, was first introduced by Hillier(12). It was then extended by Allsop to networks of a general structure(13). The method was later formulated in terms of Dynamic Programming and applied in conjunction with a computationally efficient network partitioning algorithm(14). The Dynamic Programming procedure for the general network is presented in this paper as a set of two GCM network operation rules that are a straightforward generalization of the CM rules for series-parallel networks. The procedure is further used as a tool in determining optimal network settings that take into account costs attributable to both the deterministic traffic flow model and its associated stochastic fluctuations.
Traffic Flow Model

In order to illustrate the key features of the traffic flow process, we consider an idealized model. The discrete nature of vehicular movement is disregarded and traffic is thought of as a continuous fluid. The following assumptions are made:

(a) All cars travel with uniform speed between adjacent intersections.
(b) Traffic flow is saturated, i.e., traffic volume at each intersection equals serving capability.

Let $i$ and $j$ denote two adjacent signalized intersections in the network, such that cars can travel from $i$ to $j$ along the link connecting them. Referring to Fig. 1 we define the following parameters:

- $g_j(r_j)$ - effective green (red) time of signal $j$
- $C = g_j + r_j$ - network common cycle time
- $\phi_{ij}$ - offset time between signals $i$ and $j$
- $t_{ij}$ - travel time from $i$ to $j$
- $f_{ij}(t)$ - instantaneous traffic flow (vehicles per unit of time)
- $F_{ij} = \frac{1}{C} \int_{0}^{C} f_{ij}(t) dt$ - average traffic flow

In the case that traffic is assumed to have a periodic arrival pattern of rectangular shape as illustrated in Fig. 2(a), it can be easily verified that the rate of delay (delay per unit of time) on link $(i,j)$, $d_{ij}(\phi_{ij})$, is

$$d_{ij}(\phi_{ij}) = \begin{cases} 
F_{ij} g_j(t_{ij} - \phi_{ij}) & \text{if } t_{ij} - g_j \leq \phi_{ij} \leq t_{ij} \\
F_{ij}(\phi_{ij} - t_{ij}) & \text{if } t_{ij} \leq \phi_{ij} \leq t_{ij} + r_j 
\end{cases}$$
and is similarly periodic with respect to $\phi_{ij}$ (see Fig. 2(b)). Examination of Fig. 1 indicates that offset variations can be confined to a single cycle time by introducing the transformation

$$O_{ij} = (\phi_{ij}) \mod C$$

Thus, $0 \leq O_{ij} < C$ and the resulting delay function $d_{ij}(O_{ij})$ is illustrated in Fig. 2(c). To approximate more closely traffic conditions, taking into account secondary flows, platoon dispersion etc., more elaborate models can be used (15,16,17). An example of an actual traffic flow pattern, that has been measured directly by detectors on the street is shown in Fig. 3(a). The link delay function associated with this pattern is obtained by applying elementary queueing relationships (Fig. 3(b)). While the discussion in this paper is confined primarily to delays, the optimization methods can be used with a more general link performance function combining costs of delays, stops, acceleration noise, or other measures of effectiveness, by using appropriate weighting factors (18,19).

**Criterion of Optimization**

The objective of the network optimization procedure is to determine signal settings (i.e., cycle time, splits and offsets) that minimize total delay. The total delay in the network, $D$, is regarded as a sum of two components:

$$D = D_d + D_s$$

The first component $D_d$, is the delay time resulting from the deterministic traffic flow model described above:

$$D_d = \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}(O_{ij})$$
FIGURE 3

(a) ASPECT CYCLE IN 2 SECOND INTERVALS

(b) LINK OFFSET (SECONDS)

FLOW (VEH/INT)

0 0.5 1.0 1.5 2.0

1 5 10 15 20 25 30 35

AVERAGE DELAY (SECONDS/VEH)

0 1.0 2.0 3.0 4.0

0 10 20 30 40 50 60 70

CYCLE LENGTH 70 SECS
EFFECTIVE GREEN 32 SECS
RELEASE RATE 1.76 VEH/INT

FIGURE 3
where \( n \) denotes the number of intersections (nodes) in the network. \( d_{ij} = 0 \) if link \((i,j)\) does not exist. For given cycle and splits this delay is a function of offsets only. The second component of delay, \( D_s \), is due to the stochastic nature of traffic flow. It is taken to be independent of the choice of offsets in the network but will be important for evaluating the best choice for cycle time.

The procedure for optimization consists of scanning a number of cycle times, usually in 10 sec. intervals not exceeding the range of 40 sec. to 120 sec. For each cycle time, splits at each node are calculated according to proportions of conflicting traffic streams(2), and offsets throughout the network are optimized by the Generalized Combination Method. The maximum number of offsets \( o_{ij} \) that can be assigned independent values in a network of \( n \) nodes is \( n-1 \) and the links across which they are defined must constitute a tree pattern, i.e., have no loops(20).

The computational procedure for minimizing \( D_d \) with respect to offsets involves the division of offsets into \( N \) equal intervals. It is convenient to consider the link delays to be a function of an integer number, say \( k \), \( k = 0,1,...,N-1 \). To simplify notation we also adopt the convention

\[
(x) \mod N = (x)_N
\]

The Combination Method

The Combination Method is a technique for determination of offsets that minimize delays in series-parallel networks(12). The method applies a network reduction sequence to yield a total delay function for the complete network that is represented by a single equivalent link(21). The optimizing offsets of the network are determined by minimizing this function. The reduction sequence is based on the following two rules:
CM 1: **Reduction of parallel links.** Where two or more links occur in parallel, joining a pair of nodes, the delay functions of the individual links are added with respect to the same offset to yield a combined delay function represented by a single link between the two nodes.

Application of this rule is illustrated in Fig. 4(a). Given \( d_{12}(i) \) and \( d_{21}(j) \), the combined delay \( D_{12}(i) \) for the equivalent link is

\[
D_{12}(i) = d_{12}(i) + d_{21}(N-i) \quad \text{for } i = 0, 1, \ldots, N-1.
\]

CM 2: **Reduction of series links.** Wherever a node is connected by two links with two other nodes, that node is eliminated and the two links are replaced by a single link. The equivalent delay function for this link is computed by minimizing the total delay for each offset between the extremities of the two links. At each step the procedure involves a search over all possible offsets between one of the extremal nodes and the common node and selecting the minimum.

Referring to Fig. 4(b), given \( d_{12}(i) \) and \( d_{23}(j) \) the delay function for the equivalent link is

\[
D_{13}(k) = \min_i (d_{12}(i) + d_{23}[(k-i)N]) \quad \text{for } k, i = 0, 1, \ldots, N-1.
\]

Since the three offsets \( i, j, k \) form a closed loop the constraining relationship among them in this case is

\[
j = (k-i) \mod N.
\]

**The Generalized Combination Method**

This method relieves the series-parallel restriction imposed on the structure of networks by the ordinary Combination Method. By generalizing
the rules stated in the preceding section it is possible to optimize networks of arbitrary layout (subject only to computational considerations).

**GCM 1:** Combination of partial networks. Delay functions that pertain to separate parts of a network and depend on offsets between the same set of nodes are added to produce an equivalent delay function for the combined parts of the network.

**GCM 2:** Elimination of interior nodes. An equivalent delay function for a partial network is calculated for all offsets between its boundary nodes (i.e., the nodes which disconnect it from the remainder of the network) by eliminating from the optimization process the offsets relative to its interior nodes. The values of the function are determined by minimizing the total delay of the partial network for all offsets between the boundary nodes. At each step the calculation is effected by searching over all possible offsets associated with the interior nodes and selecting the minimum.

Recursive application of these rules defines a total delay function for the complete network, for offsets between a certain final set of nodes. Optimizing offsets are determined by minimization of this function. Application of the method is illustrated in the following two examples.

**Example 1:** The network to be optimized is illustrated in Fig. 5(a). Series-parallel combination produces the v-Y configuration shown in Fig. 5(b), that cannot be further reduced. At this stage the network is disconnected into two parts and a delay function is calculated for each (Fig. 5(c)). Following GCM 2,
FIGURE 5
\[ D_1(h,i) = \min_k \{ d_{25}(k) + d_{75}((k-h)N) + d_{54}((h+i-k)N) \} \]

This partial minimization also yields the relation \( k^*(h,i) \) where \( k^* \) is the optimizing value of offset \( k \) for each combination of \( h \) and \( i \).

Following GCM 1 we obtain,

\[ D_2(h,i) = d_{27}(h) + d_{74}(i) + d_{24}((h+i)N) \]

and the total delay \( D \),

\[ D(h,i) = D_1(h,i) + D_2(h,i) \]

Minimization of \( D(h,i) \) determines the optimizing offsets \( h^* \) and \( i^* \). Backtrack computation yields the optimizing offsets for all links of the original network.

**Example 2:** The original signal network is shown in Fig. 6(a). After series-parallel reductions the compressed network of Fig. 6(b) is obtained. Optimizing offsets are calculated by stagewise partitioning of this network and recursive application of the GCM rules at each stage. A partitioning plan that minimizes number of operations and storage requirements for this network is given in this table:

<table>
<thead>
<tr>
<th>Stage Number</th>
<th>Disconnecting Nodes</th>
<th>Eliminated Interior Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2,3,4</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>5,3,4</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>5,6,4</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5,6,7</td>
<td>4,8</td>
</tr>
</tbody>
</table>

The detailed minimization process is outlined below and illustrated in Fig. 6(c).
FIGURE 6
The delay function obtained at stage 4 represents total delay in the network for each possible combination of offsets n and r. The terminal optimization stage consists of minimizing this function with respect to n and r and calculation, by backtracking, of an independent set of optimal offsets (in this case offsets $j^*, k^*, m^*, q^*, n^*, r^*, s^*$).

The Network Cycle Time

The traffic flow pattern on a signalized link can be regarded as the combination of a periodic component imposed by the preceding signal and a random component arising from variations in driving speeds, marginal friction, and turns. The latter component causes additional delay owing to the occurrence of an overflow queue at the signal's stop line. While this effect is negligible at low degrees of saturation, its importance becomes predominant at high values (22, 23, 24).

At each node in the network we have the relation

$$\sum_j g_j = C - L$$

i.e., the sum of effective green times on all phases equals the net green
time available for movement through the intersection (cycle time less lost

time). Rearranging we obtain

\[ \sum_{j} \lambda_j = 1 - \frac{L}{C} \]

where \( \lambda_j = \frac{g_j}{C} \) denotes the green split \( j \) (fraction of cycle time allotted to

phase \( j \)). The split, in turn, is determined as follows:

\[ \lambda_j = \frac{y_j}{Y} (1 - \frac{L}{C}) \]

\( y_j = \frac{F_{ij}}{S_j} \) is the representative ratio of flow to saturation flow of a particular

phase and \( Y = \sum_{j} y_j \) is the sum of \( y \)-values over all phases of the intersection.

The \( y \)-values depend only on flow and saturation flow, but not on the signal

settings themselves. The total lost time \( L \) is usually a fixed quantity at a

particular intersection. Therefore, a change in \( C \) alters the total net green
time available for passage through the intersection and, consequently, its

allotment to the phases—-the green splits. This eventually brings about a

change in the degree of saturation and with it in the size of the overflow

queue.

An estimate of the expected overflow queue, based on the capacity

of the signal's approach and the degree of saturation was given by Wormleighton.

He considered the traffic behavior along the link as a non-homogeneous Poisson

process with a periodic intensity function (25). A typical relationship

between overflow queue and split is shown in Fig. 7. Denote the expected

overflow queue at the phase with split \( \lambda_j \) by \( Q_o(\lambda_j) \). The network wide

expected delay, \( D_S \), associated with these waiting vehicles is:
\[ D_s = \sum_j Q_0(\lambda_j) \]

where the sum is to be taken over all phases at all nodes of the network. This gives the second component of the network objective function, \( D = D_d + D_s \).

Recursive application of the GCM for different cycle times, taking into account both deterministic and stochastic effects, produces typical results as shown in Fig. 8. These curves were calculated for the network shown in Fig. 6(a). It should be noted that input links must be also included in the calculation. While they do not affect signal coordination (i.e., calculation of offsets), they play an important role in evaluating the total delay for selecting the cycle time. It is evident that the optimal cycle time for the network constitutes an equilibrium point between delays caused by deterministic effects and delays caused by stochastic effects. While the first delays usually increase with cycle length, the latter decrease with it owing to the decrease in the degree of saturation (or the load factor). They approach asymptotically from above the minimal cycle time for the network, which is the theoretical minimal cycle time for the most heavily loaded intersection if all flows were deterministic. These characteristics are in complete analogy with the behavior of delay with respect to cycle time at a single intersection (2). However, the results are different and a single intersection analysis would virtually never give the optimum cycle time for the network.
SUMMARY

Traffic signal settings in a network involve cycle time, splits and offsets. The methods currently used ignore some of the important factors that affect the choice of an optimal set of settings. In this paper determination of settings in a synchronized signal network is effected by scanning a number of cycle times. For each cycle splits are determined locally at each intersection according to proportions of traffic loads on conflicting approaches. Offsets throughout the network are optimized by the Generalized Combination Method which constitutes the basic building-block of the optimization procedure.

The additional costs incurred by travellers through the network owing to the signalized intersections are regarded as a sum of two components: deterministic and stochastic. These components of costs are attributable to corresponding components of the traffic flow model. For given cycle time and splits, the deterministic costs component is a function of the offsets in the network and generally increases with cycle length. The stochastic costs component is determined by the size of the expected overflow queue at each signal-controlled approach which is a function of the capacity at the approach and the degree of saturation. This component is taken to be independent of the offsets and decreases with decreasing rate when cycle time increases. It is shown that, with this objective function optimal signal settings in a network, including the cycle time and the set of splits and offsets associated with it, are determined as an equilibrium point of minimum costs resulting from the combined consideration of the two components.
References


LIST OF FIGURES

Fig. 1: Link and signal parameters.
Fig. 2: Idealized traffic flow pattern and link delay functions.
Fig. 3: Actual platoon profile and link delay function.
Fig. 4: Series-parallel network reduction rules.
Fig. 5: Signal network for Example 1.
Fig. 6: Signal network and optimization sequence for Example 2.
Fig. 7: Overflow queue vs. green time split.
Fig. 8: Variation of total network delays with cycle time.