MICROSCOPIC ANALYSIS OF TRAFFIC FLOW PATTERNS FOR MINIMIZING DELAY ON SIGNAL-CONTROLLED LINKS

Nathan Gartner*, Massachusetts Institute of Technology

An optimal scheme for coordination of consecutive signals along arterial routes or networks requires a microscopic analysis of the traffic flow patterns on every link of the system. Such an analysis was carried out for two-way links on a major artery in downtown Toronto. Accurate platoon profiles were obtained via the digital computer system controlling traffic lights throughout the metropolitan area and its associated vehicle-detector system. Individual link delay functions were calculated subject to the particular characteristics of each signalized traffic link. These functions were then combined in parallel according to the principles of the British TRRL combination method. The optimal settings derived are shown to deviate substantially from those established by conventional coordination methods. The resultant improvement in delay to traffic was confirmed by direct field observations.

A PAIR of adjacent signalized intersections along an urban arterial street is connected by two-way traffic links. The signals are synchronized, i.e., they operate with a common cycle. The common cycle time is determined by the requirements of the most heavily loaded intersection along the arterial. The green-time splits at each intersection are determined by local demand so that equal degrees of saturation on the conflicting approaches are achieved. The pattern of traffic flow on each of the links is assumed to remain steady during the time period considered. No network constraints are imposed on the phasing of the signals under consideration. The problem is to determine the relative phase (offset) between the linked pair of signals that will cause minimal delay to traffic on the connecting links.

This is a subproblem of the more general problem concerning optimal coordination of linear signal systems. The traditional tool of the traffic engineer in coordinating signals along urban arterials has been the progressive signal system (1, 2, 3). Its objective is to maintain maximal throughbands proportionate to the traffic volumes along both directions of the artery. This objective is intuitively associated with the reduction of stops and delays to traffic. However, it has long been recognized that this design maximizes essentially a geometric quantity (the bandwidth) without explicitly taking into account the actual traffic flow patterns using the system (4, 5).

An advance on this method has been the combination method (6, 7), originally conceived by the British Transport and Road Research Laboratory (TRRL). The salient feature of this technique is that delay incurred by traffic is taken under direct consideration and is systematically minimized. The greatest difficulty in applying the technique is to determine the basic delay-offset relationship for each link. A simplified model for calculation or a crude simulation procedure is usually employed (8).

It is the purpose of this paper to describe a study in which a microscopic analysis of the actual traffic flow patterns on the signal-controlled links is undertaken. The analysis is carried out by direct measurement of traffic flows via vehicle loop detectors.

*The research reported was performed while the author was a research fellow at the University of Toronto.
Publication of this paper sponsored by Committee on Traffic Control Devices.
relating their data to the computer control center of Metropolitan Toronto. The data are then processed by the computer to obtain accurate delay-offset relationships for the links under consideration, based on the characteristics of these links. Optimal settings for the controlling signals are derived subsequently.

THEORETICAL BACKGROUND

Link Delay Functions

A traffic link is defined as a section of street carrying traffic in one direction between two signalized intersections, as shown in Figure 1. Delay is incurred at the downstream signal of the link, i.e., where traffic exits the link. The offset across any link is defined as the time difference between the starting point of a green phase at the upstream signal of the link and the starting point of the next green phase at the downstream signal. It is a directional quantity, assuming the direction of traffic flow along that link. This section describes the transition process of traffic through the link's exit signal and the computational procedure for obtaining a delay-offset relationship, given the cyclic flow pattern on the link.

For the purpose of the present discussion, a zero value is assigned to the beginning of the green time at the exit signal of the link in order to establish it as a reference point. Thus, the time interval \((-r, g)\) consists of an effective red period \((-r, 0)\) and an effective green period \((0, g)\) so that

\[ r + g = C \] (1)

where \(C\) denotes the cycle time. The following notations are also used:

- \(q_a(t)\) = arrival rate (vehicles/second),
- \(q_d(t)\) = departure rate (vehicles/second),
- \(A(t)\) = cumulative number of arrivals,
- \(D(t)\) = cumulative number of departures,
- \(s(t)\) = possible departure rate (should there be a nonexhaustive lineup of cars at the signal's stop line), and
- \(s_0\) = saturation flow rate during green period.

Starting with the beginning of any red period at the exit signal, we have the following basic relations:

\[ A(t) = \int_{-r}^{t} q_a(\tau) d\tau \] (2)
\[ D(t) = \int_{-r}^{t} q_d(\tau) d\tau \] (3)
\[ s(t) = \begin{cases} 0 & \text{if } -r < t \leq 0 \\ s_0 & \text{if } 0 < t \leq g \end{cases} \] (4)

The following assumptions are made:

1. Arrivals are periodic, i.e., for any integer number \(n\),
   \[ q_a(t) = q_a(t - nC) \] (5)

2. The signal is undersaturated, i.e.,
   \[ A_p < g s_0 \] (6)

where the total number of cars arriving during one cycle (the platoon size) is
\[ A_g = -\int_{-r}^{g} q_a(t)\,dt \]  

3. The arrival rate during the green time of the signal does not exceed the saturation flow rate,

\[ q_a(t) < s_e \quad \text{if } 0 < t \leq g \]  

This implies that, once a queue has vanished during the green period, it cannot rebuild before the next red period commences.

According to these assumptions, all vehicles arriving during a cycle in which the red period precedes the green can be accommodated in that cycle. It follows that the queue is always empty at the end of the green period, and delay time calculations can be confined to a single interval \((-r, g)\).

The queue length \(Q(t)\) at any time \(-r < t < g\) is given by the difference between the cumulative number of arrivals and the cumulative number of departures.

\[
Q(t) = \begin{cases} 
A(t) & \text{if } -r < t \leq 0 \\
A(t) - ts_e & \text{if } 0 < t \leq t_0 \\
0 & \text{if } t_0 < t \leq g
\end{cases}
\]

\(t_0\) denotes the time when the queue disappears \((0 < t_0 < g)\). By definition, \(t = t_0\) when

\[ Q(t_0) = A(t_0) - ts_e = 0 \]

If we follow this analysis, the departure rate is described by

\[
q_d(t) = \begin{cases} 
0 & \text{if } -r < t \leq 0 \\
s_e & \text{if } 0 < t \leq t_0 \\
q_a(t) & \text{if } t_0 < t \leq g
\end{cases}
\]

An illustration of the traffic transition process through the link’s exit signal is shown in Figure 2.

The delay incurred by \(Q(t)\) queuing vehicles during an interval \(dt\) is \(Q(t)\,dt\). Therefore, the total delay time \(\delta\) incurred by traffic during a complete cycle \((-r, g)\) is represented by the area under the queue-length curve.

\[
d(\theta) = \int_{-r}^{g} Q(t)\,dt = \int_{-r}^{t_0} Q(t)\,dt
\]

Obviously, the size of this area depends on the relative phase of the exit signal, i.e., on the offset \(\theta\). The average delay per car (per cycle) \(\delta(\theta)\) is obtained by dividing by the total number of arrivals during one cycle.

\[
\delta(\theta) = \frac{1}{A_p} d(\theta)
\]

The procedure described yields only one point on the delay-offset curve. To obtain the complete relationship requires that this procedure be repeated while the relative phasing between the exit signal settings and the arrivals is altered so that all possible offsets across the link under consideration are examined.

Parallel Combination Procedure

According to the principles of the combination method, where two or more links occur in parallel, joining two nodes, the delay functions of the individual links can be
combined, with reference to the same offset, to yield a total delay function. Referring to Figure 1, $d_1 (\theta_{11})$ and $d_2 (\theta_{11})$ are calculated for $0 \leq \theta_{11} \leq C$ and $0 \leq \theta_{11} \leq C$ respectively. The two offset variables in this case are constrained by the following relationship:

$$\theta_{11} + \theta_{11} = C$$

Consequently, only one of the two offsets can be determined independently. Relating the total delay $D$ to offset $\theta_{11}$, we obtain

$$D(\theta_{11}) = d_1 (\theta_{11}) + d_2 (C - \theta_{11})$$

To obtain the average combined delay function $\Delta(\theta_{11})$ from the individual average delay functions, we use the following formula:

$$\Delta(\theta_{11}) = \frac{1}{A_{p1} + A_{p2}} [A_{p1} \delta_1 (\theta_{11}) + A_{p2} \delta_2 (C - \theta_{11})]$$

An optimal offset $\theta^*_1$, between the adjacent pair of signals, is readily obtainable by searching for the minimal value of the combined function.

**SYSTEM CHARACTERISTICS**

**Links**

The site where the study was carried out is located on Bloor Street, one of the major east-west arterials in downtown Toronto. The two links selected for analysis connect the junctions of Bloor-Yonge and Bloor-Church. The links consist of two traffic lanes in each direction. Geometrical details are shown in Figure 3. Parking and stopping are prohibited during rush-hour periods, whereas, otherwise, stopping alone is permitted. All turning movements are prohibited at the Bloor-Yonge intersection. Only turning-in movements from Church Street onto Bloor Street are allowed at the Bloor-Church intersection. Thus, relatively compact platoons are formed, characterized by a high degree of coherence.

To calculate the expected delay incurred by traffic at the signal stop lines requires that an estimate of the saturation flows on these links be made. Inasmuch as traffic is confined to well-marked lanes, the Australian method for capacity calculations was adopted (9, 10). The results in through car units (tcu's) are given in Table 1.

**Detectors**

In the Toronto signal system, vehicle detectors are installed on all approaches to signals that have been designated to operate in the TR2 responsive control mode (11). Under this scheme green-time splits are adapted dynamically to traffic demand, while a constant cycle length and a fixed offset relationship in the major flow direction are maintained. Both signals at the study site shown in Figure 3 operate in the TR2 control mode and thus are equipped with detectors. Each link has one 4-ft wide magnetic loop detector across both traffic lanes, located approximately 300 ft in advance of the downstream intersection. Information on vehicular presence within each detection zone is transmitted via telephone lines to the traffic control center by means of tone-telemetering devices. At the center, the state of each detector circuit is scanned at a rate of 16 times a second. The data are then fed into a Univac 418 computer to be reduced and relayed to a Univac 1107 computer for further processing.

**DATA COLLECTION AND ANALYSIS**

**Platoon Profiles**

The traffic data reported by the detectors to the central computer are used for calculations in the on-line responsive green-time allocation algorithm. They are also
Figure 1. Two-way traffic links connecting pair of adjacent signalized intersections.

Figure 2. Traffic transition process through a link's exit signal. (U and V in arriving and departing flow patterns are equal in magnitude. W represents the estimated delay per cycle.)

Figure 3. Geometrics of study site: link 1—eastbound; link 2—westbound.

Figure 4. Platoon profiles and delay functions for 70-sec cycle length: (a) link 1 and (b) link 2.

Table 1. Saturation flows by link and lane.

<table>
<thead>
<tr>
<th>Link</th>
<th>Curb Lane (tcu/hour)</th>
<th>Center Lane (tcu/hour)</th>
<th>Total for Approach (tcu/hour)</th>
<th>(tcu per 2-sec interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (eastbound)</td>
<td>1,270</td>
<td>1,580</td>
<td>2,850</td>
<td>1.56</td>
</tr>
<tr>
<td>2 (westbound)</td>
<td>1,350</td>
<td>1,650</td>
<td>3,000</td>
<td>1.67</td>
</tr>
</tbody>
</table>
recorded on magnetic tape in conjunction with the prevailing signal aspects for off-line analysis and evaluation. Thus, if the signal upstream from a detector is being controlled with a fixed cycle length, the platoon profile at the detector can be calculated from the data recorded by the computer. The cycle is divided into short subintervals of time (usually 2 sec). The number of vehicles crossing the detector in each interval is averaged over a number of cycles during which the average flow intensity is assumed to remain steady (e.g., 25 cycles). The resulting curve shows how the vehicle density varies throughout the average cycle. Congested cycles, i.e., cycles during which the approach queue on the downstream end of the link extends to the detector or beyond, must be excluded; otherwise, the actual shape of the platoon profile will be distorted.

The method used to detect congestion is based on an exponential smoothing process. Average values are continuously generated by assigning relative weights to new versus old information according to the following generalized formula: (present value of average) = A (new data) + (1 - A) (previous value of average). The smoothing factor A in this formula has to be a positive number less than 1. Exponential smoothing is used to calculate both the average volume level EDGEX and the average pulse length PULSEX. These values are calculated for each approach as follows:

\[
\text{EDGEX}_n = (\alpha) \text{count}_n + (1 - \alpha) \text{EDGEX}_{n-1}
\]
\[
\text{PULSEX}_n = (\beta) \text{pulse}_n + (1 - \beta) \text{PULSEX}_{n-1}
\]

\(\alpha\) and \(\beta\) are smoothing factors that have been assigned the values 0.01 and 0.2 respectively. Count\(_n\) denotes the number of counts recorded by the detector during the time interval \(m\). Pulse\(_n\) denotes the \(n^{th}\) pulse length, i.e., the time duration for which the induction loop senses vehicle presence; it is inversely related to speed. The PULSEX calculation is performed only when a new pulse occurs, whereas the EDGEX calculation is repeated for each time interval. Congestion is assumed to exist when both of the smoothed quantities pass specified threshold values. The values are chosen so that only the existence of a real queue is detected, while disturbances such as single vehicles stopping or moving slowly over the detector are ignored. The values used by the traffic program are 0.07 for EDGEX and 0.7 for PULSEX (12).

The analysis for each link is terminated when the required number of cycles has accumulated or if a program change has occurred and the cycle time of the upstream signal was altered. In this case the flow pattern over the detector changes, and a new profile has to be generated.

Shift and Calibration Procedures

The platoon profile obtained through averaging of the magnetic tape vehicle count recordings, which was described in the previous section, has to undergo two transformations, one in time and the other in magnitude.

Time Transformation—The profile describes the arrival pattern observed at the detector location during an average cycle starting with the upstream signal's green phase. If the link delay function is to be calculated, this pattern has to be extrapolated (in time) to the stop line at the downstream signal. With well-defined platoons, such as those shown in Figures 4 and 5, travel times from the previous intersection to the detector can be estimated. The point at which the platoon profile rises steeply represents the average arrival time of the leading edge of the platoon. Given the length of the link and if allowances are made for the fact that the leading edge of the platoon usually accelerates from a standing position, a good estimate of speed can be made. The average speed that was detected on the links under consideration varied in the range of 25 to 28 mph. The arrival pattern is then shifted backward by the expected travel time from the detector to the downstream stop line.

Magnitude Transformation—The two factors involved are the necessity for count calibration due to errors in detector countings and the conversion of vehicle counts into equivalent tcu's. Extensive field observations have shown that, due to a multitude of error possibilities, the detectors tend to overcount for volumes below 600 vph per
two-lane approach, whereas they undercount for volumes above this level (13). A calibration curve obtained by regression analysis is available to take account of these errors. Because the saturation flows are specified in tcu/hour, the arrival flow has to be stated in the same units. If the average composition of traffic is known as well as the proportions of turning vehicles, the counts can be scaled up to equivalent tcu's by using appropriate conversion factors.

Random Variations

Once the platoon profile is obtained, the delay-offset relationship can be readily established by repeated use of Eqs. 1 to 16. However, this profile represents the expected pattern of arrivals at the signal’s stop line. It is an average component of a periodic process that includes also a random component arising from variations in driving speeds, marginal friction, and turns. The latter component may cause additional delay because of the possible occurrence of an overflow queue $Q_o$ at the end of the green period. Although this effect is negligible at low degrees of saturation, its importance increases at values exceeding 0.80 (14). According to our notation (see Eq. 6), the degree of saturation $x$ is defined as

$$x = A_s / g_s$$

(17)

Account of this factor is taken by estimating the expected overflow queue according to the value of the saturation flow at the signal’s approach and the expected degree of saturation. Such an estimate was given by Wormleighton (15), who considered traffic behavior along the link as a nonhomogeneous Poisson process with a periodic intensity function represented by the platoon profile obtained from the detector countings.

Results

Data were collected during the morning rush hours while an inbound plan (70-sec cycle) and a heavy inbound plan (80-sec cycle) were in operation. The resultant platoon profiles for each of the links involved are shown in Figures 4 and 5 respectively. The ratio of flows among the links is approximately 2:1 in favor of the inbound direction (westbound direction in this case). The conventional procedure in such circumstances is to provide the best possible progression to accommodate the heavy flow. In some cases this includes provision of advance clearance times for queues that might develop during the signal's red period owing to overflows, turning movements, or intralink sources (such as parking lots and garages) and obstruct the major flow emanating from the previous signal. Given the length of the link $l$ and the desired speed of travel along the link $v$, the desired offset between the signals in that direction, $\theta_s$, would have to be

$$\theta_s = \frac{l}{v} - \frac{Q_s}{s_o}$$

(18)

$Q_s$ denotes the expected queue length at the start of downstream green and combines the overflow queue $Q_o$ contributed by the random fluctuations and the expected number of arrivals during the red period: $\int_{-T}^{0} q_s \, dt$. The value of $Q_s$ is determined either by estimation or by direct observation.

According to this design, the starting of downstream green is advanced $Q_s / s_o$ sec prior to the regular progression time along the link of $l/v$ sec, so that the platoon released from the upstream signal will be able to pass unhindered. In fact, the advance time might have to be larger to account for possible additional arrivals joining the queue while it is being dissolved. The basic philosophy underlying this design is one of smooth flow control. However, no direct quantitative analysis in terms of delays or other costs is attempted to support this supposition. Two deficiencies are apparent:

1. The coordination policy is concerned only with the leading edge of the arriving platoon, while the remainder of it is neglected; and
2. Traffic on the opposing link (on a two-way arterial), which carries a much lighter load, is often disregarded.

The effects of these deficiencies are illustrated in the following analysis.

Based on the observed platoon profiles, a delay-offset relationship is calculated for each link. Each point on the curve represents the average delay per vehicle the platoon would incur should the downstream green start at the corresponding time of the arriving pattern. Because the origin coincides with the starting of upstream green, the horizontal axis describes the relative phase between the signals (the offset). The individual link delay functions are depicted in Figure 6 for the 70-sec cycle plan and in Figure 7 for the 80-sec cycle plan. For each of these pairs a combined relation is established from Eq. 16 and is shown illustrated in Figures 6c and 7c respectively. A range of offsets for which the expected combined delay is within 5 percent of the minimum has been designated as the "minimum range." Point A indicates the offset advocated by the progression method according to Eq. 18. Evidently, a considerable reduction in delay time, and hence in travel time, can be achieved by selecting one of the offsets within the minimum range: up to 40 percent for the 70-sec cycle plan and 45 percent for the 80-sec cycle plan.

Field Observations

To check the validity of the link delay functions, we conducted several field observations at the test site. The performance of the traffic signal control system was compared for different offset settings while an "inbound" plan (70-sec cycle) was in operation. The comparison was done by manual measurement of stopped-time delays. A detailed description of the observations is given elsewhere (16). The compiled results of measurements for offsets corresponding to points A and B in Figure 7c are given in Table 2.

Westbound Link (link 2)—This link carries the major flow during the morning rush hours. While traveling along the link, the platoon formed by the upstream signal disperses in time. As it arrives at the downstream signal, it has spread wider than the available green service period, and some of the cars must be delayed. With an offset corresponding to point A, the leading edge of the platoon arrives just after the queuing cars have cleared the approach (or at the beginning of green should there be no waiting cars) and can pass unimpeded. However, a substantial portion of the platoon is cut off when the green time terminates and has to wait for the next phase to be served. As was shown in a theoretical study by Newell (17), under such conditions delay in one-directional traffic is minimized if the trailing edge of the platoon arrives at the light just before it turns red and suffers no delay, whereas the leading edge arrives too early and is stopped; it is released at saturation flow when the next green commences. Obviously, an offset setting corresponding to point B comes close to this policy.

Eastbound Link (link 1)—This link carries the minor flow during the morning rush hours. The signals on this link operate at a much lower degree of saturation (approximately 0.4 as opposed to 0.8 on the westbound link). The width of the arriving platoon does not usually exceed the available green time. With an offset located at point A, the platoon arrives at the downstream signal shortly after the beginning of red and consequently suffers a long waiting time and a high percentage of stoppages. When the offset is shifted to point B, the beginning of downstream green is advanced (the offset for this link is shortened). The platoon still arrives during the red aspect but at a later instant and therefore suffers a shorter waiting time before being served. The proportion of stopped cars is also reduced slightly. Minimal delay for travel on that link cannot be achieved because of the constraint imposed on its phasing by the requirements of the parallel link, carrying traffic in the opposing direction between the same pair of signals. However, the minimum range of offsets in the combined delay function is largely due to the possibility of trade-offs in the allocation of delays among the competing links.
Figure 5. Platoon profiles and delay functions for 80-sec cycle length: (a) link 1 and (b) link 2.

Figure 6. Delay per vehicle on (a) link 1, (b) link 2, and (c) links 1 and 2.
Figure 7. Delay per vehicle for (a) link 1, (b) link 2, and (c) links 1 and 2.

![Graph showing delay per vehicle for links and links 1 and 2.]

Table 2. Comparison of coordination policies by field observations of stopped-time delays.

<table>
<thead>
<tr>
<th>Item</th>
<th>Point A</th>
<th>Point B</th>
<th>Percentage of Change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Link 1</td>
<td>Link 2</td>
<td>Links 1 and 2</td>
</tr>
<tr>
<td>Average delay per vehicle (sec)</td>
<td>12.6</td>
<td>16.1</td>
<td>14.9</td>
</tr>
<tr>
<td>Percentage of vehicles stopped</td>
<td>85.0</td>
<td>53.8</td>
<td>63.9</td>
</tr>
<tr>
<td>Volume (vph)</td>
<td>542</td>
<td>1,100</td>
<td>1,624</td>
</tr>
</tbody>
</table>
CONCLUSIONS

A microscopic analysis of traffic flow patterns was conducted using the facilities of the Toronto computer control system. It was shown that optimal signal settings determined by such an analysis can produce considerable savings in delays to traffic. As a result, travel times through the system are reduced, and the capacity of the system is increased.

The method of analysis in this study is applied to fixed-time signal settings. It is assumed implicitly that flow patterns are constant over a certain control period, such as morning or evening rush hours, off-peak periods, and the like. It is also assumed that these patterns are of a repetitive nature during similar days of the week. Experience with the Toronto system over a number of years, where extensive data collection and analysis have been carried out, has shown that these assumptions are valid for many links of the system. Hence, this procedure can be a useful tool in providing optimal coordination schemes both for arterial streets and for networks.

The outlined method is not limited to fixed-time programs. Conceptually, the same procedure can be applied to develop a responsive control logic by which optimal settings are determined for control periods on the order of a few minutes. Apart from the hardware requirements, the main problem is whether programs should be selected automatically according to demand or calculated on-line.

An interesting subject for further investigation would be to assess the importance of other criteria in traffic signal coordination. In some schemes, a stop penalty is superimposed on the delay cost (18). As the limited field tests conducted in this study have shown, reduction in delays was accompanied by a reduction in the number of stops. But, generally, the minima of both measures do not necessarily coincide, though they are close (19). Emphasis is placed, in these cases, on the deterministic act of a complete stop, i.e., the deceleration of a vehicle to zero velocity. The process of vehicle platoons traversing a series of traffic lights is one in which repeated decelerations and accelerations take place. Some of the vehicles have to decelerate to a complete stop. A rational objective function should take into account the disutilities associated with all vehicular maneuvers. These disutilities might eventually include the costs of time losses, wear, discomfort, accident risk, pollution, etc. The delay function proposed in the present analysis is believed to constitute a good approximation to such a generalized cost function.

ACKNOWLEDGMENT

The author gratefully acknowledges the assistance given to him by E. Hauer and R. M. Soberman of the University of Toronto in conducting this research. The author would also like to thank S. Cass, Metropolitan Toronto Commissioner of Roads and Traffic, for making it possible to use the facilities of the Toronto Traffic Control Centre and for his permission to publish the results.

REFERENCES