portation Networks. Academic Press, New York, 1972.
9. M. L. Manheim. Fundamentals of Transportation Systems Analysis. Cambridge, Mass., in press.
10. F. V. Webster. Traffic Signal Settings. British Road Research Laboratory, Her Majesty's Stationery Office, London, BRRL Rept. 39, 1961.
11. N. Gartner. Area Traffic Control and Network Equilibrium. In International Symposium on Traffic Equilibrium Methods (M. Florian, ed.), Springer-Verlag, New York, 1976.
12. F. A. Wagner, F. C. Barnes, and D. L. Gerlough. Improved Criteria for Traffic Signal Systems in Urban Networks. NCHRP, Rept. 124, 1971.
13. L. Rach and others. Traffic Signal Control Strategies—A State-of-the-Art. Toronto Department of Roads and Traffic, Mar. 1974.
14. N. H. Gartner, J. D. C. Little, and H. Gabbay. Simultaneous Optimization of Offsets, Splits and Cycle Time. TRB, Transportation Research Record 596, 1976, pp. 6-15.
15. M. J. Maher and R. Akcelik. The Redistributional Effects of an Area Traffic Control Policy. Traffic Engineering and Control, Sept. 1975.
16. N. H. Gartner. Analysis and Control of Transportation Networks by Frank-Wolfe Decomposition. Proc., 7th International Symposium on Transportation and Traffic Theory, Kyoto, Japan, Aug. 1977.

Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.

Abridgment

Optimum Control of Traffic Signals at Congested Intersections

Panos G. Michalopoulos, Rensselaer Polytechnic Institute, Troy, New York

Optimum control of congested intersections continues to be one of the major concerns of traffic engineers. Several theories have been proposed for and successfully applied to optimum control of uncongested intersections, but relatively little theoretical work has been applied to congested or oversaturated conditions. Thus, at present we must rely on practical evidence to time signals at oversaturated intersections.

To be sure, the applicability of existing theories is severely limited because they require complex instrumentation and extensive computations or because the assumptions made in deriving the policy are oversimplified. Further, none of the proposed policies incorporates both of the two necessary criteria for optimum control of oversaturated intersections: (a) minimizing total intersection delay and (b) subjecting (a) to queue length constraints. I therefore conclude that the dependence of current practice on empirical considerations is not entirely unjustified. It can also be shown (1) that some of the practical control policies are in fact best applied under specific conditions.

In order to promote a better understanding of optimum control strategy, I shall first present an isolated congested intersection. Then, I shall extend the theory to a system of two or more intersections.

ISOLATED INTERSECTIONS

Before developing the stragegy, I need to further clarify the first criterion for optimum operation. When minimizing intersection delay at congested intersections, one should consider total delay for the entire oversaturation period rather than delay per cycle. This is so because a per cycle optimum does not, in general, correspond to the optimum policy for the entire peak period. Wrong conclusions, therefore, might be drawn.

The optimum control policy for minimizing total intersection delay was first developed by Gazis and Potts and D'Ans (2, 3, 4, 5). However, the second and in some instances most important objective of the control was not incorporated in the process. Thus, an analytical solution to the problem when queue length constraints are present is still lacking. However, I have adopted Gazis' approach here because it provides a reasonable basis for complete formulation of and solution to the problem.

The oversaturated intersection can be formulated and treated as a control problem in which: (a) the system is the intersection, (b) minimization of the aggregate delay subject to queue length constraints is the objective of the control, (c) the queues on each approach to the intersection describe the state of the system, and (d) cycle length and splits are bounded control variables.

Derivation of the optimum control policy is obtained by first considering the simplest case: an intersection of two one-way streets. It should be noted here that extending this method to more complex situations is straightforward and should not present any problems. The demands on each approach vary with time as in reality and are known for the entire control period. If we assume that demand becomes higher than capacity at roughly the same time on both approaches, that the cycle is constant, and that variable green times are bounded, we can formulate the problem by inspecting the cumulative input-output curves on each approach. The area between these two curves represents the total delay that must be minimized, and the problem is formally stated as follows: Minimize the delay function

$$\text{Min } D = X_0 = \int_0^T (x_1 + x_2)dt$$

and subject it to

$$f_0 = dx_0/dt = x_1 + x_2$$

$$f_1 = dx_1/dt = q_1(t) - u$$

$$f_2 = dx_2/dt = q_2(t) - s_2(1 - L/c) + (s_2/s_1)u$$

$$0 \leq i \leq \alpha_i$$

with the boundary conditions

$$x_i(0) = x_i(T) = 0$$

$$x_0(0) = 0$$

and the admissible control domain

$$s_1 g_{min}^{(1)}/c = u_{min} \leq u = s_1 g_1/c \leq s_1 g_{max}^{(1)}/c = u_{max}$$

where D represents total delay, T is the end of the oversaturation period, and c is the cycle length. The remaining parameters are detailed as follows:

$x_i(t)$ = the queue length on approach i (and i = 1, 2);
$q_i(t)$ = the input flow rate on approach i;
$u(t)$ = the control variable = $[s_1 g_1(t)]/c$;
α_i = the upper bound placed on queue i;
s_i = the saturation flow on approach i, which is assumed constant;
$g_i(t)$ = the effective green time on approach i as defined by Webster (6);
$g_{max}^{(1)}(g_{max}^{(1)})$ = the maximum (minimum) green time allowed on approach i; and
L = the total lost time per phase.

Space limitations preclude presentation of the details here, but the major results are summarized below (1).

1. When no queue length constraints are imposed, the optimum control policy is essentially the bang-bang control suggested by Gazis (7, p. 213), which, in more concrete traffic terms, implies that the best policy comprises two stages. During the first stage maximum green time should be given to the approach with the highest saturation flow. During the second stage, the operation should be switched so that minimum green time is given to the approach with the maximum saturation flow, maximum green time to the other approach.

2. When queue length constraints are imposed, several switch points may exist. Then the control is not necessarily bang-bang, because it can be shown (1) that the control variable u (and therefore the green times) may receive intermediate values. More specifically, the optimum control is bang-bang as long as none of the queues reaches its upper or lower bound. If this does happen, the optimum control is to switch the signals (to change u) as follows:

a. When the queue on the first approach reaches its upper or lower bound, the optimum control u* is

$$u^*(t) = q_1(t) \tag{1}$$

b. When the second queue reaches either of its bounds, the control is given by

$$u^*(t) = [s_2(1 - L/c) - q_2]s_1/s_2 \tag{2}$$

3. Examination of the above relationships leads to the conclusion that, when both queues reach their upper or lower bounds during overlapping time intervals, cycle length must then be free to vary according to the relationship

$$c = L/1 - [(q_1/s_1) + (q_2/s_2)] \tag{3}$$

4. This implies that when two constraints are imposed, the optimum solution requires a variable cycle in addition to variable splits. If, however, only one constraint is imposed, the cycle can remain constant. Thus an upper and a lower bound should be placed on the cycle for optimum control with queue length constraints.

5. We can now see that when the intersection is not saturated the optimum control policy is to switch the signals as soon as the queue receiving green is exhausted. This policy has, in fact, been proposed by Dunne and Potts (8) and is known as the saturation flow algorithm. To be sure, the operation of traffic-actuated signals is based on this principle. It has been well established from experience that traffic-actuated control results in minimum delays at isolated intersections.

The major problem associated with the optimum control strategy is determining the switch points (if they exist), the final time, and the resulting minimum total delay. Because of the complexity of the problem, one must be content with a numerical solution generated by a digital computer when nonlinearities are introduced (nonlinear time-varying demands). It should be noted that a closed form solution requires analytical expressions describing the history of arrivals that vary from intersection to intersection; this makes the numerical approach even more attractive. I developed an eight-step algorithm allowing the determination of the optimum control policy for a given situation and present examples demonstrating the applicability and the effectiveness of the theory (1).

SYSTEMS OF INTERSECTIONS

In extending the theory to a system of two or more oversaturated intersections, we must take travel times and queuing storage between intersections as well as turning movements into account. In order to simplify the demonstration of this theory, consider a system of two oversaturated intersections of three one-way streets. Making assumptions similar to those for the single intersection and taking into consideration the average travel time d_1 between the two intersections, one can formulate the problem following the guidelines as before. However, the resulting optimum control problem is not trivial because there are two control variables, and the one associated with the upstream intersection appears in the process both without delay and with delay d_1. Space limitations preclude presentation of the details here, but a summary of the optimum control policy to minimize total system delay subject to queue length constraints shows:

1. The entire control period (O, T) should be divided into two intervals (O, T - d_1) and (T - d_1, T), where T represents the end of the control period (or the final time) defined as the time at which the last queue is gone.

2. During both intervals, the resulting optimum control is bang-bang with at most three switchover points per intersection, assuming that none of the queues reaches its upper or lower bound.

3. As soon as the queue between the two intersections reaches its upper or lower bound, the control action taken at the downstream intersection becomes an explicit function of the control action taken at the upstream intersection d_1 time units earlier. This simply means that for optimum control the two intersections must be coordinated. Coordination is more clearly demonstrated for the special case in which turns at the upstream intersection are prohibited. Thus, if $u_1(t)$ and

$u_2(t)$ represent the control action taken at the upstream and the downstream intersections respectively, it can be shown (1) that for optimum control $u_2(t) = u_1)(t - d_1)$, which indicates that in this special case the control action at the downstream intersection should be identical to the control action taken at the upstream intersection d_1 time units earlier.

4. For the remaining queues results similar to those presented for the single intersection can be obtained, because the former are not affected by the time delay.

5. The cycle at each intersection can remain constant as long as no more than one queue per intersection exceeds its bounds at a time. Otherwise, as expected, optimum control becomes more complex, involving variable cycles and splits as the single intersection does.

Extension of the theory to more than two intersections does not present any problem because it is a straightforward application of the pair of intersections. It should be noted, however, that, as the number of intersections increases or when a large number of queue length constraints are imposed, the optimum control policy becomes more complex. In the final analysis, then, it would be too difficult to apply it to a pretimed system. I therefore suggest that, as system and performance requirements increase, closed loop control (computer control) should be considered.

CONCLUDING REMARKS

The theory developed here clearly provides an analytical solution to the problem of optimum control of traffic signals at congested intersections with queue length constraints. The theory can also be employed for computerized control, because initial and final conditions (queues) can be other than zero. It should be pointed out, however, that in this case adaptive prediction algorithms are needed in order to change the control policy as new information concerning the state of the system and the predicted demands arrives. Such algorithms have been developed and used by computer control schemes, but it is generally agreed (9) that further research in this area is needed.

Examples demonstrating the applicability and effectiveness of the theory to real-life situations can be found in another of my works (1), where it is clearly shown that the optimum control, compared with conventional control techniques, can result in substantial delay savings. In these examples, as expected, queue length constraints resulted in a more complex optimum control policy. Further, the fixed time policy was still more effective than conventional control even for demands substantially different from the predicted levels. It is perhaps of interest to note that determining the optimum control strategy requires only knowledge of conventional traffic parameters such as saturation flow; 5-, 10-, or 15-min counts; and maximum allowable green intervals.

To repeat, extending the theory not only should not present any problem, it can also include a large-scale traffic signal network with both critical and uncongested intersections. By using principles of graph and control theory along with eigenvalue analysis, determination of "dominant" intersections (not necessarily oversaturated), defined as those requiring immediate control action at a given instant, can be made. Subsequently, optimum control action will be taken at these intersections alone, eliminating extensive data manipulation and computer computations. This approach should result in instantaneous and efficient decisions made by the computer and time-varying rather than fixed critical intersections. Dominant intersections can be identified by eigenvalue analysis because they are functions of the demand and the control decisions made in the past.

Other problems associated with oversaturated intersections, such as optimum cycle length determination and bus priorities at such intersections, have been studied, and results will be presented in the near future.

REFERENCES

1. P. G. Michalopoulos. Optimal Control of Oversaturated Signal Systems: Some Theoretical Considerations. Univ. of Florida, Gainesville, PhD thesis, 1975.
2. D. C. Gazis and R. B. Potts. The Oversaturated Intersection. Proc. 2nd International Symposium on the Theory of Road Traffic Flow, Organization for Economic Co-operation and Development, Paris, 1965, pp. 221-237.
3. D. C. Gazis. Optimum Control of a System of Oversaturated Intersections. Operations Research, No. 12, 1964, pp. 815-831.
4. D. C. Gazis. Modelling and Optimal Control of Congested Transportation Systems. Networks, No. 4, 1974, pp. 113-124.
5. G. C. D'Ans and D. C. Gazis. Optimal Control of Oversaturated Store-and-Forward Transportation Networks. Transportation Science, Vol. 4, No. 1, 1976, pp. 1-19.
6. F. V. Webster. Traffic Signal Settings. British Road Research Laboratory, Her Majesty's Stationery Office, London, BRRL Rept. 39, 1958.
7. D. C. Gazis. Traffic Control—Theory and Application. In Traffic Science (D.C. Gazis, ed.), Wiley, New York, 1974, pp. 175-237.
8. M. C. Dunne and R. B. Potts. Algorithm for Traffic Control. Operations Research, No. 12, 1964, pp. 870-881.
9. J. B. Kreer. Factors Affecting the Relative Performance of Traffic Responsive and Time of Day Traffic Signal Control. Transportation Research, No. 10, 1976, pp. 75-81.

Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.