Influence of Control Measures on Traffic Equilibrium in an Urban Highway Network

Nathan H. Gartner, Federal Highway Administration, U.S. Department of Transportation

Traffic control measures play a critical role in determining equilibrium between demand and supply in an urban highway network. Present techniques attempt to optimize network performance assuming a fixed demand pattern. Experience indicates that control measures can be used to influence demand patterns in such a way that total network performance is significantly improved. On the basis of the interdependence of control measures and resulting traffic flow patterns, this paper develops an analytical framework for a systematic optimization of the operation of a highway system. Two practical examples demonstrate such an optimization on a limited scale, and a model and a mathematical program for achieving it are presented.

Across the nation close to 150 cities and municipalities have turned to computerized traffic control systems to relieve congestion, to save time and energy, and to reduce pollution (1). We expect that within the next decade every urban area of 100 000 or more will have a computer system of one kind or another controlling traffic flow on its streets and freeways (2, 3). These systems are bound to have a profound impact on the performance of transportation networks. However, the electronic hardware seems to be far more advanced than practicing traffic engineers are yet capable of utilizing. It is the purpose of this paper to describe a framework for systematically optimizing the operation of such systems within the broader context of traffic equilibrium in a network.

The effects of various traffic engineering measures on traffic flow in urban areas has been the subject of many investigations, several of them sponsored by the National Cooperative Highway Research Program, Area 3: Traffic—Operations and Control (4).

One particular project on optimizing flow on existing street networks (5) stated its objective as developing "a practical method of measuring the degree of change in network traffic flow resulting from various system modifications." This objective was based on the recognized need to better understand the effects that many commonly used transportation system modifications have on traffic flow. The measures studied included directional control and lane use, curb lane controls, channelization, signal controls, and bus operation. The conclusions of this study suggested that, in improving a downtown area, those elements that involve the functional use of streets, such as one-way patterns, reversible lane operations, and major parking prohibitions, should be developed first. Then all the minor influences that create friction in the traffic stream, such as turning movements, truck loading, pedestrian interference, proper allocation of signal time to the various approaches of an intersection, and other traffic problems of the area, should be analyzed and corrected. Only when local friction has been adequately reduced can traffic be properly platooned to implement signal progressions. After this has been successfully accomplished, the best locations for bus stops may be considered, and bus movement in the progressive system may be developed.

There is no doubt that traffic flow can significantly benefit from such operational improvements in an urban street network. In most cases, the benefits to the traveling public will considerably outweigh the costs of analysis, engineering, and construction of the improvement. However, most of these improvements tend to reinforce the existing traffic patterns rather than to explore alternatives to them (6). This is most evident in calculating and updating signal control settings, which is based upon the existing or projected pattern of traffic. The result is that signal controls are regularly set to favor existing heavy flows of traffic over light flows. This is particularly evident when arterial progression schemes rather than areawide optimization schemes are being implemented. Consequently, the routes that drivers already use and the destinations they most frequently select are made more attractive, and the ones they usually reject are made even less attractive. Thus, the choices that created existing traffic flow patterns receive constant reinforcement.

Many traffic management schemes indicate that it may be desirable, for one reason or another, to reduce the amount of traffic using parts of the road network (7). It is quite conceivable that, in concert with other control measures, this could be achieved by means of traffic signals. Signals would be installed at the points of entry to the area concerned, and the settings would be used to regulate flow into the area. But even without restraining vehicle movements or forcing changes of destination, a change in the traffic patterns may be beneficial simply because drivers select a better route.

It is well known that individual route choices do not generally lead to the least total travel costs for all users of the system (8). This is so because drivers do not choose routes to enhance the common good but to minimize their travel costs. The resulting flow pattern (termed "user-optimal") is not necessarily the best for the community (termed "system-optimal"). Therefore we must decide whether traffic controls can be used to influence individual route choice in order to achieve a traffic flow pattern that benefits all. Two examples described in this paper indicate that this is indeed possible.

TRAFFIC EQUILIBRIUM

The core of any transportation analysis problem is predicting flows that will use all the segments of the system. The pattern of traffic flows in a road network is the result of choices subject to various constraints made by a large number of individuals. Such constraints may be of an economic nature, such as the income level of the individual, or technological, such as the performance characteristics of the available modes of transportation. Each individual, given the particular constraints that apply, chooses whether to travel at all or if so to which destination, at which time, by which mode, and by which route. The collective choices made by all individuals in a transportation system will determine a set of flows and levels of service that can be viewed as an equilibrium condition between the demand for and the supply of trans-

portation services. The basic hypothesis underlying this statement is that there is a market for transportation that can be separated from other markets (9).

The basic paradigm of equilibrium in a transportation network is illustrated in Figure 1A. Let us define

- L = level of service (such as trip time) on a particular facility;
- V = volume of flow on this facility;
- S = supply;
- D = demand;
- T = specification of the transportation system in terms of its technological characteristics (including operating policies); and
- A = specification of the socioeconomic activity system.

Then the supply function $L = S(T, V)$ shows an increase in the level of service as volume increases, and the demand function $V = D(A, L)$ a decrease in volume as the level of service increases (in the negative sense). The resulting equilibrium point $E_o(L_o, V_o)$ occurs at the intersection of the two curves.

In computing traffic equilibrium for an urban street network with signalized intersections, one usually assumes that the origin-destination demand pattern is an inelastic function of the level of service. This is a reasonable assumption when estimating short-run equilibrium, that is, when no trip distribution changes are expected during the anlysis period. The interdependence between signal settings and flow patterns can also be explained with reference to the equilibrium paradigm shown in Figure 1B. A change in signal settings causes the entire supply curve to shift. On the other hand, a change in flow corresponds to a movement along the curve. For example, a change from curve S_1 to S_2 would result in a decrease in the level of service measure from L_1 to L_2 for the same flow V_1, whereas an increase in flow from V_1 to V_3, perhaps from increased supply, would cause an increase in the level of service measure from L_2 to L_3. The following sections examine the interaction of control measures and traffic volumes in determining equilibrium conditions in an urban road network.

TRAFFIC CONTROL

Intersections are an important source of delay in urban areas, and many are controlled by traffic signals. Therefore, analysis of signalized intersections is of primary importance in determining traffic equilibrium in an urban network. The average delay (d) per vehicle on a signalized approach can be represented, according to Webster (10), as the sum of two components

$$d = d_s + d_d = 0.9\left(\left\{[C(1-g)^2]/[2(1-q/s)]\right\} + \left\{(3600x^2)/[2q(1-x)]\right\}\right) \quad (1)$$

where

- g = G/C, proportion of green time (G) to cycle time (C);
- q = arrival flow (vehicles/h);
- s = saturation flow (vehicles/h);
- K = gs, capacity of the approach (vehicles/h);
- x = q/K, degree of saturation;
- d_d = delay if flow is deterministically uniform; and
- d_s = additional delay caused by stochastic nature of traffic flow.

The way delay varies with traffic flow (or, equivalently, with the degree of saturation) is illustrated in Figure 2. The characteristics closely resemble the transportation supply functions shown in Figure 1. Both green time and cycle time must be determined, and all flows must be considered before we can derive the level of service (delay) at which traffic through the intersection will be served. Analyses of single intersection controls can be found in the literature (10, 11).

When two or more intersections are in close proximity, some form of linking is necessary to reduce delays and to prevent frequent stopping. Because signalized intersections have a platooning effect on leaving traffic, it is advantageous to synchronize signals to operate within a common cycle. It also becomes necessary to coordinate the signals, that is, to establish an offset between the signals so that loss to traffic is minimized. Networks, like single intersections, also require that all demands be considered simultaneously when control variables are being determined.

Several computer methods have been developed for optimizing signal settings in networks (12, 13). One method particularly suitable for the application discussed in this paper is the mixed-integer traffic optimization method (MITROP) (14). In this, for each link (i, j) in the network a link performance function comprising a deterministic and a stochastic component is constructed. The deterministic component represents the average delay incurred per vehicle in a periodic flow through signal j. The stochastic component arises from variations in driving speeds, marginal friction, and turns and is expressed by the occurrence of overflow queues at intersection stop lines. Thus, we can also consider the total delay in the network to be composed of two components, D_d and D_s. The network signal setting problem can then be stated in a general form as

1. Determining offsets ϕ_{ij}, splits G_{ij}, cycle time C;
2. Minimizing total delay $D = D_d + D_s$; and
3. Subjecting them to loop offset constraints $\sum_{loop} \phi_{ij} = n_\ell C$ (n_ℓ - integer), to link split contraints $G_{ij} + R_{ij} = C$, and to signal capacity constraints $q_{ij}C \le s_{ij}G_{ij}$.

This is a nonlinear mathematical program with integer decision variables and can be solved by appropriate computer optimization packages.

TRAFFIC CONTROL AND ROUTE CHOICE

Traditionally traffic engineers have chosen signal settings that minimize travel costs by using any of the methods mentioned above, given a fixed traffic flow pattern. It is assumed that all route choices are predetermined and result in constant flows on each link, irrespective of the controls imposed on that link and, hence, irrespective of the level of service offered by the link. This assumption is, in most cases, incorrect. As shown in Figure 3, there exists an interdependence between traffic controls and link volumes; one determines the other and vice versa. We must therefore develop a model that incorporates both traffic controls and route choice (traffic assignment) and thus provides a tool for establishing mutally consistent signal settings and traffic flow patterns. The following two examples are given to illustrate quantitatively the object of such a model.

Example 1

Figure 4A shows the intersection of Commonwealth

Figure 1. Basic equilibrium paradigms in a transportation network.

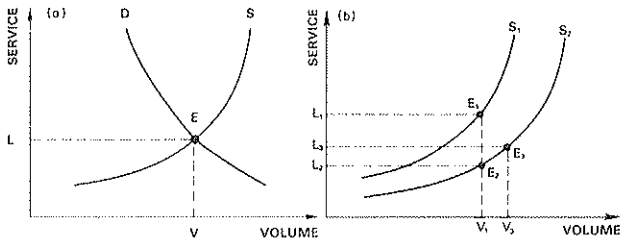


Figure 2. Average delay to traffic on a signal controlled intersection approach.

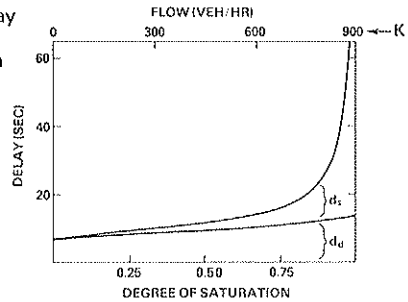
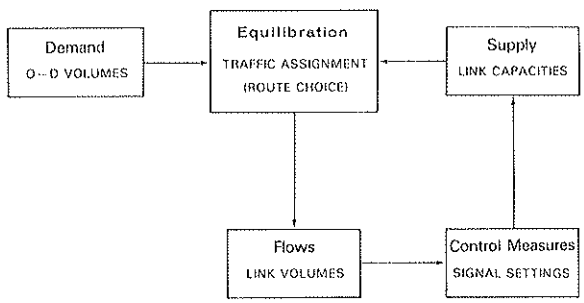


Figure 3. Interdependence of link volumes and traffic controls.



Avenue at Boston University Bridge when it has an advanced green phase allowing west-to-north left turns. The average delay for the left turners is 63.0 s, and the rate of total delay for all traffic passing through the intersection is 41.6 vehicles/s^2 (Table 1).

Turning to Figure 4B, it is seen that west-to-north movements can also be accomplished by traveling an extra 75 s through route $A_1 \rightarrow E \rightarrow F \rightarrow B_2$. The signal can now be operated on a shorter cycle time, because the left-turn movement is prohibited. Total expected travel time on the extended route is 22.0 + 75.0 + 16.7 = 103.7 s, an increase of 64 percent (Table 2). However, the total rate of delay for all users of the intersection is now only 32.3 vehicles/s^2, a reduction of 22.3 percent with respect to the previous figure.

It should be noted that despite the fact that a left-turn arrow is no longer shown, it is physically possible but illegal to make the turn against the opposing traffic. Occasionally vehicles do indeed make the turn. But, taking into account the good chance of getting a costly ticket, which can be regarded as a very long delay, most drivers will choose the longer route, thus achieving system optimization at this location.

Example 2

This example is based on results reported by Maher and Akcelik (15). Signal settings and traffic flows were calculated iteratively by the following procedure:

1. Optimizing signal settings for the current flow patterns;
2. Reassigning origin-destination demands corresponding to the prevailing signal settings by using a capacity restraint technique; and
3. Iterating among steps 1 and 2 until no significant changes occur in successive link flow patterns and signal settings.

The particular network studied is described in Figure 5 and this origin-destination matrix:

From	To X	To Y
A	600	600
B	0	600
C	0	2000

Two flow patterns are calculated, following the two principles enunciated by Wardrop. One is the following user-optimized flow pattern, representing the way traffic will distribute itself among alternative routes.

Flow	Route 1 to 3	Route 2 to 3
A to Y	0	600
B to Y	180	420

Total network travel time in this user-optimizing pattern is 60.8 vehicles/h^2. The second is this system-optimized flow pattern, representing the way traffic should distribute itself in order to minimize travel time for all users of the system.

Flow	Route 1 to 3	Route 2 to 3
A to Y	0	600
B to Y	600	0

Total network travel time in this system-optimizing pattern is 53.7 vehicles/h^2. It is seen that the second flow pattern produced a reduction of 12 percent in travel time and could be achieved simply by banning left-turns for flow B to Y at intersection 1.

Discussion

These examples clearly show the advantages of combining traffic control and route choice. The problem is how to incorporate route choice as part of the traffic control optimizing program. One approach is based on the general formulation given by $D = D_d + D_s$ and $q_{ij}C \leq s_{ij}G_{ij}$. Only origin-destination trip demands are given, and all link flows q_{ij} are decision variables. The objective function is modified to also include the travel time on the links T, in addition to the delay at the signals,

$$T = \sum_{i,j} q_{ij} t_{ij} \tag{2}$$

where t_{ij} is the travel time on link i, j. A new set of constraint equations is added to represent the orgin-destination demands and continuity of flow by nodes. In matrix notation it takes the following form

$$\underline{A} \times \underline{q} = \underline{p} \tag{3}$$

where $\underline{A}$ is the node-link incidence matrix of the network; $\underline{q}$ is the link flow vector; and $\underline{p}$ is the origin-destination demand vector.

The solution to this program represents a system-optimized flow pattern and control program. By modify-

Figure 4. Signalization schemes for example 1.

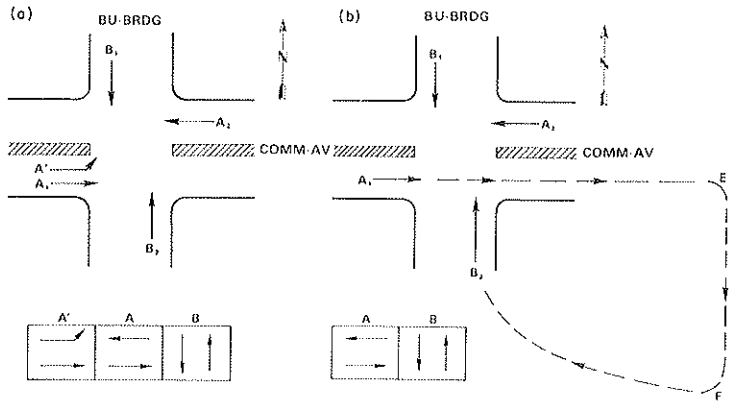


Table 1. Delays for traffic at three-phase approach intersection.

Approach	Arrival Flow (vehicles/h)	Saturation Flow (vehicles/h)	Green Time/ Cycle Time[a] (cycles)	Degree of Saturation	Average Delay (s)	Total Rate of Delay[b] (vehicles/s²)
A′	240	1500	0.18	0.880	63.0	4.2
A_1	1500	3000	0.54	0.925	30.7	12.8
A_2	1200	4500	0.36	0.750	28.8	9.6
B_1	1050	3000	0.38	0.920	42.0	12.2
B_2	500	3000	0.38	0.440	20.4	2.8

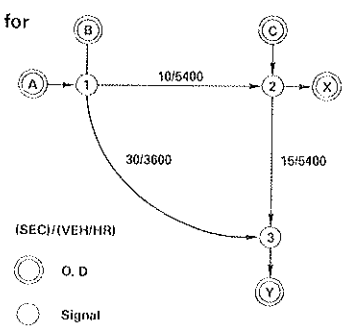
[a]Cycle = 100 s. [b]Level of service = 9 s.

Table 2. Delays for traffic at two-phase approach intersection.

Approach	Arrival Flow (vehicles/h)	Saturation Flow (vehicles/h)	Green Time/ Cycle Time[a] (cycles)	Degree of Saturation	Average Delay (s)	Total Rate of Delay[b] (vehicles/s²)
A_1	1740	4500	0.46	0.85	22.0	10.6
A_2	1200	4500	0.46	0.59	16.0	6.8
B_1	1050	3000	0.43	0.815	22.6	6.5
B_2	740	3000	0.43	0.58	16.7	3.4

[a]Cycle = 100 s. [b]Level of service = 9 s.

Figure 5. Test network for example 2.



ing the objective function, the same program can be used to determine a user-optimized flow pattern. As shown in example 2, comparisons of the two patterns should prove instructive to the traffic engineer in deciding which control procedures and management techniques are helpful in improving traffic flow. Methods of solving such a program can be developed by using modern traffic assignment techniques, such as those based on Frank-Wolfe decomposition (16).

CONCLUSIONS

Traffic equilibrium in an urban road network is the result both of control measures taken by the traffic engineer and of choices made by the individual driver. The examples described in this paper demonstrate that determining control measures consistent with the traffic flow patterns is preferable to calculating control programs and flow patterns independently of each other. The nonlinear optimization program, based on the MITROP optimization model, was presented to achieve this objective on a networkwide basis.

REFERENCES

1. Computerized Traffic Control Systems. Traffic Engineering, Vol. 45, No. 4, Apr. 1975.
2. J. F. Schlaefli. Street Traffic Control by Computer—A State-of-the-Art Discussion. Workshop on Computer Traffic Control, Univ. of Minnesota, Mar. 1974.
3. N. R. Kleinfield. Computerized Traffic Control. The Wall Street Journal, Aug. 13, 1974.
4. Summary of Progress through 1976. NCHRP, 1976.
5. W. E. Pontier, P. W. Miller, and W. H. Kraft. Optimizing Flow on Existing Street Networks. NCHRP, Rept. 113, 1971.
6. R. E. Allsop. Use of Traffic Signals to Influence the Amount and Routing of Traffic. Conference on Getting the Most From Our Transport Facilities—The Role of Traffic Engineering. U.K. Transport and Road Research Laboratory, Crowthorne, England, Apr. 1975.
7. W. D. Berg and P. J. Tarnoff. Research in Urban Traffic Management. TRB, Transportation Research Record 603, 1976, pp. 55-56.
8. R. B. Potts and R. M. Oliver. Flows in Trans-

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