OPERATIONS RESEARCH CENTER

working paper



MASSACHUSETTS INSTITUTE OF TECHNOLOGY

AREA TRAFFIC CONTROL AND NETWORK EQUILIBRIUM

Ьy

Nathan H. Gartner

OR 041-75

J

March 1975

Supported in part by the U.S. Army Research Office (Durham) under Contract No. DAHC04-73-C-0032

ABSTRACT

Area traffic control systems play an important role in determining the equilibrium between demand and supply in an urban highway network. The paper describes some of the methods that have been developed for the operation of these systems and derives the level-of-service that would result from a given set of flows. Currently used techniques attempt to optimize network performance assuming a fixed pattern of demands. It is shown, via an example, that control measures can be used to affect the demand pattern in such a way that total network performance is improved. A model for achieving this objective is discussed.

1. INTRODUCTION

According to a recent survey⁽¹⁾ more than 100 urban areas in the United States and Canada are in the process of planning, designing and implementing area-wide computerized traffic control systems. It is anticipated that within the next few years every city of more than 100,000 people will have such systems in operation for controlling the traffic flow through its intersections and freeway corridors.⁽²⁾ These systems are bound to have a profound impact on the performance of the traffic network in those areas in terms of level-of-service and capacity. It is the purpose of this paper to describe some of the methods that have been developed for the operation of these systems and to discuss possible further developments.

The methods that are used for setting signals to control the flow of traffic through the intersections of a highway network are, perhaps, the most practical application of the concepts of traffic equilibrium in a transportation network. The systems in which they are implemented have the additional capability of monitoring the traffic flow, keeping track of its time-varying dynamics in great detail via vehicle detectors, and responding quickly to changes in demand via the communication system linking detectors, signals, and computer altogether. Unfortunately, the electronic hardware seems to be far more advanced than the engineering knowledge is capable of making use of it.

The basic paradigm of equilibrium in a transportation network is illustrated in Fig. 1A. $^{(3)}$ Let us define:

- L = level-of-service (such as trip time) on a particular facility
- V = volume of flows on this facility
- T = specification of the transportation system (including its control measures)

1.

A = specification of the activity system

Then the supply function

$$L = S(T, V)$$
 (1.1)

shows an increase in the level-of-service as volume increases, and the demand function

$$V = D(A, L)$$
 (1.2)

a decrease in volume as the level-of-service increases (in the negative sense). The resulting equilibrium point $E(L_o, V_o)$ occurs at the intersection of the two curves.

In the context of computing the traffic equilibrium in a signal-controlled highway network the simplifying assumption is made that demand is an inelastic function, fixed at a flow pattern F_o . This assumption is a good approximation when estimating short-run equilibrium. The equilibrium values in this case, $E(L_o, F_o)$, represent the level-of-service at which the given demand is serviced (Fig. 1B). It is also useful in calculating long-run equilibrium, such as in the traffic assignment phase of the transportation planning process, and has been used widely in previous studies. ^(4,5) Models for determining this equilibrium are described in the following sections.

2. THE CONTROLLED INTERSECTION

The most important element determining the level-of-service of traffic in an urban area is the at-grade signal-controlled intersection. The effect of traffic flow on travel time between intersections is usually minor compared to its effect on the delay time incurred at the intersection itself.⁽⁶⁾ Therefore, the primary determinant of the level-of-service variable L becomes the delay time. This section describes traffic performance at a single isolated signal-controlled intersection.

Let us consider first one approach to a signalled intersection. Assuming that arriving traffic is not modulated by any nearby controlling devices, the average delay per vehicle on the approach, d, can be regarded as the sum of two components

$$d = d_s + d_d \tag{2.1}$$

where d_d is the delay that would result if the flow were uniform and d_s is the additional delay caused by the stochastic nature of traffic flow. Let us define,

- C = the signal cycle time (sec)
- G = effective green time for the approach (sec)
- g = G/C, proportion of cycle which is effectively green
- q = arrival flow on approach (veh/sec)
- s = saturation flow at the signal stop line (veh/sec)
- x = q/gs, degree of saturation; ratio of flow to maximum possible flow under given setting.

According to Webster, (7) the average delay per vehicle on an approach can be approximated by

$$d = K\left[\frac{C(1 - g)^2}{2(1 - q/s)} + \frac{x^2}{2q(1 - x)}\right]$$
(2.2)

where K is a constant (usually 0.9). The predictions made by this formula were found to correlate very well with field data. Similar expressions were also derived by $Miller^{(8)}$ and Newell.⁽⁹⁾

The way the delay varies with traffic flow (or, with the degree of saturation) is illustrated, for a typical case, in Fig. 2. It is seen that at high degrees of saturation the delay rises steeply. Theoretically, the delay increases to infinity as the flow approaches capacity (x + 1.0). But in practice the flow does not sustain a high value for a long period; it falls off at the end of the peak period, and the queue does not reach a length required to cause excessive long delays. Furthermore, it has been observed that drivers, upon seeing a long queue-up away from the intersection, will turn off and seek alternative routes.

To derive the level-of-service (i.e., the delay) at which traffic through the intersection will be served, both the <u>green times</u> G and the <u>cycle time</u> C have to be determined and all flows must be considered. Let us examine, for example, the signalled intersection analysed in Fig. 3. The signal has two phases corresponding to the two possible directions of movement, N - S and E - W. The sum of the effective green times for the phases is

$$G_{EW} + G_{NS} = C - L \qquad (2.3)$$

where L, in this case, is the total 'lost time' for the intersection. Using equ. (2.2) to calculate the average delay per vehicle on each approach, and noting the symmetry in the flows, we obtain for the rate of delay, D, for each phase,

$$D_{EW} = (q_E + q_W)d_{EW}$$

$$D_{NS} = (q_N + q_S)d_{NS}$$
(2.4)

The rate of total delay, D, considering all flows through the intersection is

$$D = D_{\rm EW} + D_{\rm NS}$$
(2.5)

The three functions in equ. (2.5) are illustrated in Fig. 3 for various combinations of effective green times satisfying equ. (2.3) and a fixed cycle time C. $(G)_{min}$ is the minimum effective green time that still can accommodate

4.

the demand on the approach, though at a very high rate of delay, and is given by

The apportioning of green time among the conflicting streams at the intersection can be formulated as the following optimization program:

$$\begin{array}{c} \underline{Given} & q_{i}, s_{i} \text{ for all approaches i, and C} \\ \underline{Find} & G_{j} \text{ for all phases j to} \\ & Min D = \sum_{j} D_{j} \\ \underline{Subject \ to} \\ & \sum_{j} G_{j} = C - L \\ & G_{j} \ge (G_{j})_{min} \end{array} \right)$$

$$(2.7)$$

The optimal solution is obtained as an equilibrium point where the marginal rate of delay for the conflicting phases is equalized. In the simple two-phase example this would be

$$\frac{\partial D_{EW}}{\partial G_{EW}} = \frac{\partial D_{NS}}{\partial G_{NS}}$$
(2.8)

Webster gives an approximate rule for determining the optimal splits of green time,

$$G_{j}^{*} = (C - L)y_{j}/Y$$
 (2.9)

where y_j is the maximum ratio of flow to saturation flow for the different approaches having simultaneous right-of-way during phase j, and

$$Y = \sum_{j} y_{j}$$
(2.10)

To determine the optimum cycle time C^* for the intersection, capacity considerations play an important role. For each approach i we must have

$$q_i C \leq s_i G_i$$
 (2.11)

When the approaches are assigned to phases j we obtain,

$$y_j \leq G_j/C$$
 (2.12)

Summation over all phases j at the intersection yields

$$\Sigma y_j \leq \Sigma g_j$$
 (2.13)

Using equs. (2.3) and (2.10), we obtain

$$Y \leq 1 - L/C$$
, or, $C \geq L/(1 - y)$ (2.14)

The minimum cycle time, C_{min}, i.e., the cycle time that will serve all phases at a degree of saturation of unity, is

$$C_{\min} = L/(1 - Y)$$
 (2.15)

Obviously, because of the randomness in arrivals, such a cycle will cause an intolerable amount of delay. When plotting the total rate of delay at an intersection, D, as a function of cycle time (splits being optimized independently for each cycle), one obtains a family of curves as shown in Fig. 4. It is seen that the rate of delay is asymptotic to the minimum cycle ordinate. It decreases toward a minimum at higher cycle times as the increased capacity causes a reduction in the stochastic delay component. At still higher cycle times the rate of delay increases again. At this stage the deterministic delay component becomes dominant because of the larger red times on each approach. Webster shows that the optimum cycle time is roughly twice the minimum cycle time, i.e.,

$$C^* \cong 2C_{\min} = 2L/(1 - Y)$$
 (2.16)

Miller⁽¹⁰⁾ derives an expression for the optimal cycle time that gives similar results. On the other hand, Allsop⁽¹¹⁾ develops an iterative procedure to determine delay-minimizing settings to all traffic passing through the intersection.

3. AREA TRAFFIC CONTROL

When two or more intersections are in close proximity, some form of linking is necessary to reduce delays to traffic and prevent frequent stopping. A signal-controlled intersection has a platooning effect on the traffic leaving it, and it is advantageous to have the signals synchronized, i.e., operating with a common cycle time. It also becomes necessary to <u>coordinate</u> the signals, i.e., to establish an <u>offset</u> between the signals, so that loss to traffic is minimized. A typical platoon profile that was measured at the downstream end of a signalized traffic link is shown in Fig. 5A. Fig. 5B shows the associated link performance function, i.e., the delay incurred by the platoon when passing through the downstream signal as a function of the offset between the upstream and downstream signals. The maximum of the function corresponds roughly to the head of the platoon arriving at start of downstream red and the minimum corresponds roughly to the tail of the platoon arriving at the end of downstream green.⁽¹²⁾

The usual procedure for setting signals on arterials and in networks involves three steps. (13) First, a common cycle time is determined according to the requirements of the most heavily loaded intersection. Then splits of green time are apportioned at each intersection according to the interacting flow/ capacity ratios. Lastly, a computer optimization procedure is used to determine a set of offsets throughout the network. Several computer programs have been developed for determining offsets in a network [e.g., references 14-21], the main differences among them being in the way the traffic flow process is modeled and the optimization technique employed. Practically all these methods have shown an improvement in network performance, when compared with the earlier manual methods that were used by traffic engineers. (22,23)

However, the three-step sequential decision process used for setting the signals is not entirely satisfactory. Similar to the single intersection case where all the interacting flows have to be considered when determining optimal cycle time and splits, the network case too requires that all demands be considered simultaneously when determining the control variables. A model for such a program is presented below.

The signal-controlled traffic network consists of a set of links (i, j) connecting the adjacent signals S_i and S_j . Let,

The link performance function is composed of a deterministic delay component and a stochastic delay component. The deterministic component, $z_{ij}(\phi_{ij}, R_{ij}, C)$, is given by the average delay incurred per vehicle in a periodic flow through S_j . The stochastic component, which arises from variations in driving speeds, marginal friction, and turns, is expressed by the occurence of an overflow queue $Q_{ij}(R_{ij}, C)$ at the stop line of S_j . An estimation of the expected overflow queue was given by Wormleighton, ⁽²⁴⁾ who considered traffic behavior along the link as a non-homogeneous Poisson process with a periodic intensity function represented by the flow pattern on the link. Therefore, we can consider the total delay in the network, D, to be composed of two components

$$D = D_d + D_s \tag{3.1}$$

where,

$$D_{d} = \sum_{i,j}^{\Sigma} q_{ij} z_{ij}(\phi_{ij}, R_{ij}, C)$$
(3.2)

$$D_{s} = \sum_{i,j}^{\Sigma} Q_{ij}(R_{ij}, C)$$
(3.3)

A number of constraint equations involving the decision variables are necessary to model the network. First, the algebraic sum of offsets around any loop of the network must equal an integral multiple of the cycle time,⁽²⁵⁾ i.e.,

$$\sum_{(i,j)\in l} \sum_{\ell=1}^{\Sigma} \phi_{ij} = n_{\ell}^{C}$$
(3.4)

where n_{ℓ} is an integer number associated with loop ℓ . Effective green and effective red are related by,

$$G_{i,i} + R_{i,i} = C$$
 (3.5)

In order for the network to be able to handle the given flow we must have for each link the capacity constraint

$$q_{ij}C \leq s_{ij}G_{ij}$$
(3.6)

Practical considerations, including pedestrian crossing times and driver behavior prescribe

$$R_{ij} \ge (R_{ij})_{min}$$
(3.7)

$$C_{\min} \leq C \leq C_{\max}$$
(3.8)

Assuming, for simplicity, two-phase intersections we also have

$$R_{ij} - 1_{ij} = G_{kj} + 1_{kj}$$
 (3.9)

where (i, j) and (k, j) are assigned conflicting phases at S_{j} .

Thus, the network signal setting problem can be stated in a general form as the following nonlinear optimization program:

Find ϕ_{ij} , R_{ij} , $C \pm 0$: Min $D = D_d + D_s$ Subject to: $(i,j) \in \ell^{\pm} ij = n_{\ell}C$ $G_{ij} + R_{ij} = C$ $R_{ij} - 1_{ij} = G_{kj} + 1_{kj}$ $q_{ij}C \leq s_{ij}G_{ij}$ $R_{ij} \geq (R_{ij})_{min}$ $C_{min} \leq C \leq C_{max}$ (3.10)

 $G_{ij}, R_{ij} \ge 0; n_{\ell}$ integer.

This program is nonlinear in the objective function and in the loop constraint equations, and has integer decision variables. By a suitable representation of the objective function, the program can be solved by mixed-integer programming. $(^{26})$ In principle, the program can also be solved by dynamic programming, though, for computational reasons the splits have to be determined independently rather than simultaneously with the other decision variables. $(^{27})$

It is important to study the sensitivity of the network objective function with respect to the network cycle time. A typical relationship is shown in Fig. 6. Each point in the graph represents network performance optimized with respect to splits and offsets at a fixed cycle time. It is evident that the optimal cycle time for the network constitutes a least-cost equilibrium point between delays occuring in a deterministic situation and delays contributed by stochastic factors. While the first component of delay usually increases with cycle length (though at a decreasing rate), the latter component decreases with it because of the increase in capacity at higher cycle times. The stochastic component, and the total delay, is asymptotic from above to the minimal cycle time for the network, which is the theoretical minimum for the most heavily loaded intersection if all flows were uniform periodic. These characteristics are analogous to those observed at a single intersection, however, the implications regarding signal settings in a network are different and a single intersection analysis would virtually never give the optimum settings for the network.

4. TRAFFIC CONTROL AND ROUTE CHOICE

The foregoing discussion has considered methods for determining traffic control settings that minimize total cost, given a fixed pattern of traffic flows. It is assumed that all route choices are fixed, resulting in constant flows on each link regardless of the controls imposed on that link and, hence, regardless of the level-of-service that is offered by the link. This assumption would be correct only in the event that the level-of-service on the controlled links is insensitive to the control settings, which is, of course, incorrect. It seems, therefore, to be of fundamental importance to have a model which incorporates both traffic controls and route choice and provides a tool for establishing a system-optimized traffic flow pattern. A practical example illustrates the argument.

11.

Fig. 7A shows the intersection of Commonwealth Avenue/Boston University Bridge in Boston with an advanced green phase allowing for west-to-north left-turns. The average delay for the left-turners is 63.0 seconds and the rate of total delay for all traffic passing through the intersection is 41.6 veh x sec/sec. Turning to Fig. 7B it is seen that west-to-north movements can also be accomplished by traveling an extra 75 seconds through route $A_1 - E - F - B_2$. The signal can now be operated on a shorter cycle time by disallowing the left-turn movement. The total expected travel time on the extended route is 22.0 + 75.0 + 16.7 = 103.7 seconds, i.e., an increase of 64% in travel time. However, the total rate of delay for all users of the intersection is now only 32.3 veh x sec/sec, i.e., a reduction of 22.3% with respect to the previous figure. It should be noted that despite the fact that no left-turn arrow exists anymore, it is physically possible to make the turn, through the opposing traffic, and occasionally some vehicles are making it. But taking into account that there is a good chance of getting a costly ticket, which can be regarded to be equivalent to a very high delay, most of the drivers will choose the longer route, thus achieving a system-optimization at this location.

The question still remains, how to incorporate route choice as part of the traffic control optimizing program (for a general discussion on this subject, see a recent paper by Allsop⁽²⁸⁾). One possible approach might be to use the general formulation given by equs. (3.10), and let the q_{ij} 's be decision variables rather than input parameters. The objective function vould have to be modified to include also the cost of traveling on the links, $\sum_{i,j} q_{ij} * t_{ij}$, where t_{ij} is the travel cost on link (i, j). And to the coni,j straints, a set of equations would have to be added, respresenting the 0 - D demands and the nodal continuity of flow equations. This set will be of the form,

(4.1)

 $\underline{\mathbf{A}} \times \underline{\mathbf{q}} = \underline{\mathbf{p}}$

where,

A = node-link incidence matrix of the network

q = link flow vector

p = nodal trip productions (attractions) vector

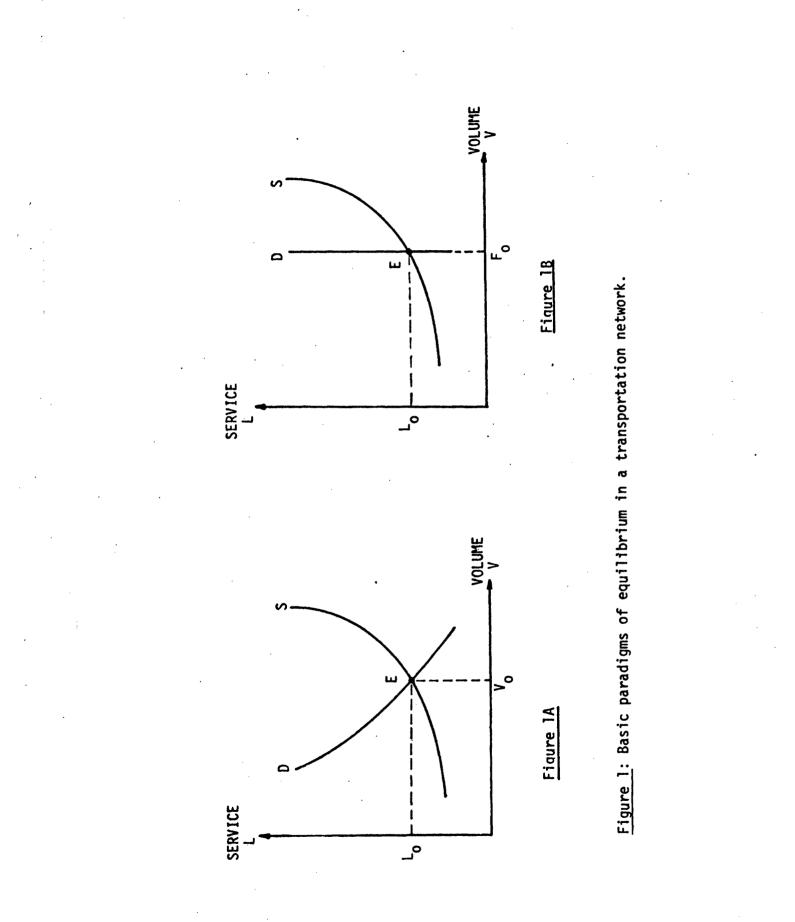
A solution to this nonlinear optimization program would represent a systemoptimized flow pattern and control program. Recent results on traffic assignment indicate that such a pattern may not be significantly different from a user-optimized pattern.⁽²⁹⁾ Alternatively, the settings obtained with this program could be used in conjunction with a suitable simulation program to arrive at a user-optimized equilibrium flow pattern that would also constitute a minimization of community costs.

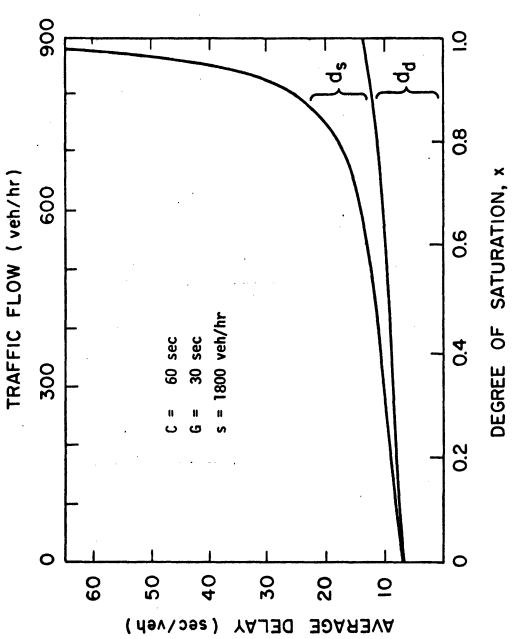
Route choice and route control are also connected with the time-varying characteristics of traffic flow, and the possibilities of exerting dynamic, traffic-responsive, control. In theory, the potential savings could be substantial. However, no successful schemes have been reported to-date, (30)and the area is open for further research.

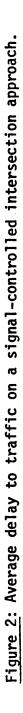
5. REFERENCES

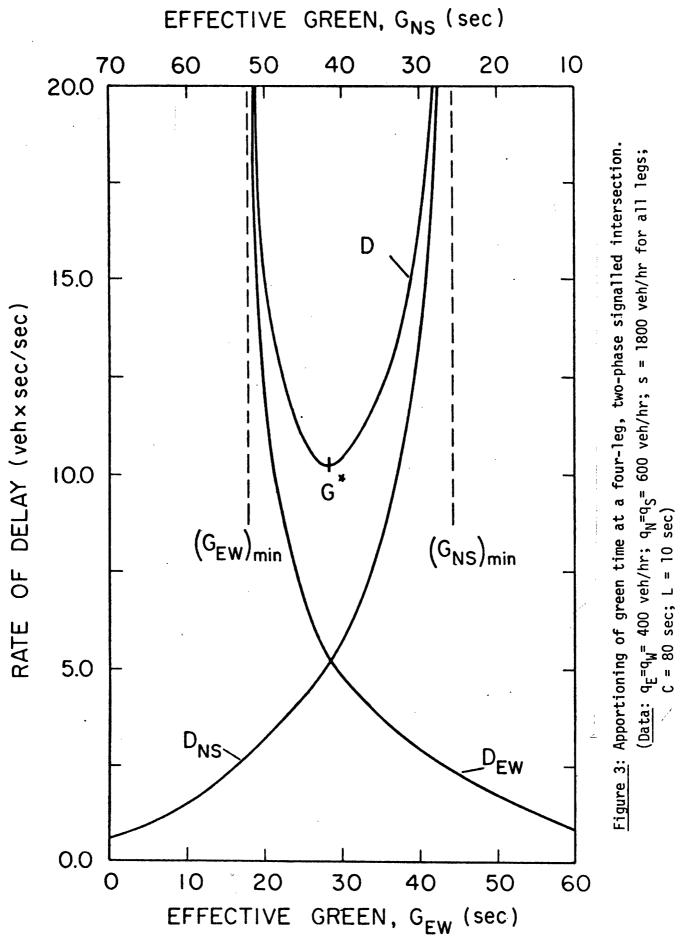
- 1. J.F. Schlaefli, "Street Traffic Control by Computer--a state-of-the-art discussion," Workshop on Computer Traffic Control, University of Minne-sota, March, 1974.
- 2. N.R. Kleinfield, "Computerized Traffic Control," The Wall Street Journal, August 13, 1974
- 3. M.L. Manheim, "Search and Choice in Transport Systems Analysis," Highway Research Record No. 293: Transportation Systems Planning, 1969.
- 4. N.A. Irwin and H.G. Von Cube, "Capacity Restraint in Multi-Travel Mode Assignment Programs," Highway Research Board Bulletin 347: Trip Characteristics and Traffic Assignment, 1962.
- 5. Comsis Corp., "Traffic Assignment--methods, applications, products," U.S. Department of Transportation, Federal Highway Administration, 1973.
- 6. "Highway Capacity Manual," Highway Research Board, Special Report 87, 1965.
- 7. F.V. Webster, "Traffic Signal Settings," Road Research Technical Paper No. 39, H.M. Stationery Office, London, 1961.
- 8. A.J. Miller, "Settings for Fixed-Cycle Traffic Signals," Operational Research Quarterly <u>14</u>, 1963.
- 9. G.F. Newell, "Approximation Methods for Queues with Application to the Fixed-Cycle Traffic Light," SIAM Review 7, 1965.
- 10. A.J. Miller, "The Capacity of Signalized Intersections in Australia," Australian Road Research Board Bulletin No. 3, 1968.
- 11. R.E. Allsop, "Delay-Minimizing Settings for Fixed-Time Traffic Signals at a Single Road Junction," J. Instr. Maths. Applics. 8, 1971.
- N. Gartner, "Microscopic Analysis of Traffic Flow Patterns for Minimizing Delay on Signal Controlled Links," Highway Research Record No. 445: Traffic Signals, 1973.
- 13. J. Holroyd, "The Practical Implementation of Combination Method and TRANSYT Programs," Transport and Road Research Lab Report LR 518, Crowthorne, 1972.
- 14. J.T. Morgan and J.D.C. Little, "Synchronizing Traffic Signals for Maximal Bandwidth," Operations Research 12, 1964.
- 15. J.D.C. Little, "The Synchronization of Traffic Signals by Mixed-Integer Linear Programming," Operations Research <u>14</u>, 1966.
- 16. Traffic Research Corp., "SIGOP: Traffic Signal Optimization Program," PB 173 738, 1966.

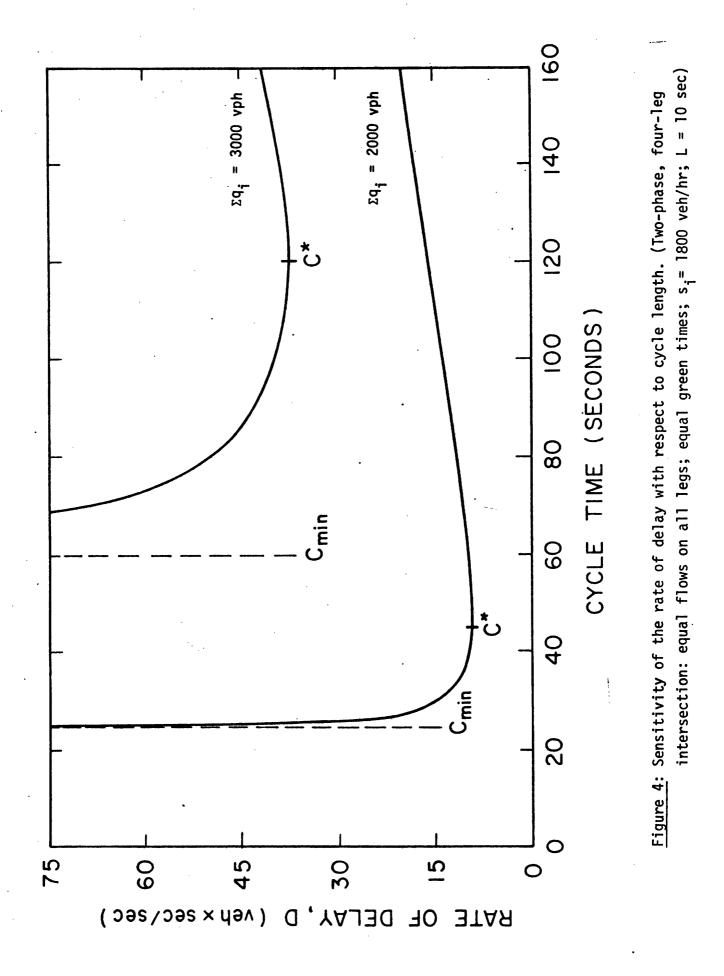
- 17. D.I. Robertson, "TRANSYT: A Traffic Network Study Tool," Road Research Laboratory Report LR 253, Crowthorne, 1969.
- 18. J.A. Hillier, and R.S. Lott, "A Method of Linking Traffic Signals to Minimize Delay," 8th International Study Week in Traffic Engineering, Theme V: Area Control of Traffic, Barcelona, 1966.
- 19. K.W. Huddart and E.D. Turner, "Traffic Signal Progressions--GLC Combination Method," Traffic Engineering and Control, 1969.
- 20. R.E. Allsop, "Choice of Offsets in Linking Traffic Signals," Traffic Engineering and Control, 1968.
- N. Gartner, "Optimal Synchronization of Traffic Signal Networks by Dynamic Programming," Traffic Flow and Transportation (G.F. Newell, Ed.), American Elsevier, New York, 1972.
- 22. J. Holroyd and J.A. Hillier, "Area Traffic Control in Glasgow: a summary of results from four control schemes," Traffic Engineering and Control, 1969.
- 23. F.A. Wagner, F.C. Barnes, and D.L. Gerlough, "Improved Criteria for Traffic Signal Systems in Urban Networks," NCHRP Report 124, Highway Research Board, 1971.
- 24. R. Wormleighton, "Queues at a Fixed Time Traffic Signal with Periodic Random Input," Canadian Operations Research Society Journal 3, 1965.
- 25. N. Gartner, "Constraining Relations Among Offsets in Synchronized Signal Networks," Transportation Science <u>6</u>, 1972.
- 26. N. Gartner, J.D.C. Little, and H. Gabbay, "Optimization of Traffic Signal Settings in Networks by Mixed-Integer Linear Programming," Technical Report No. 91, Operations Research Center, M.I.T., Cambridge, 1974.
- 27. N. Gartner and J.D.C. Little, "The Generalized Combination Method for Area Traffic Control," Transportation Research Record (to appear).
- 28. R.E. Allsop, "Some Possibilities for Using Traffic Control to Influence Trip Distribution and Route Choice," Proceedings 6th Intern. Symp. on Transportation and Traffic Theory, Sydney, 1974.
- 29. L.J. Leblanc and E.K. Morlok, "An Analysis and Comparison of Behavioral Assumptions in Traffic Assignment," Intern. Symp. on Traffic Equilibrium Methods, Montreal, 1974.
- 30. J. Holroyd and D.I. Robertson, "Strategies for Area Traffic Control Systems: Present and Future," Transport and Road Research Lab. Report LR 569, Crowthorne, 1973.











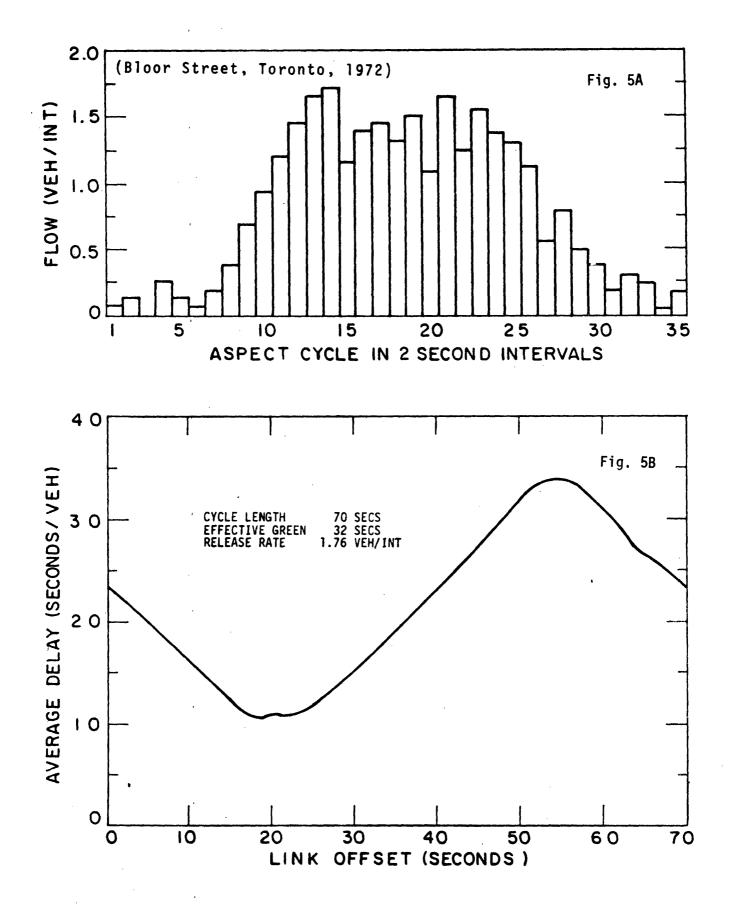
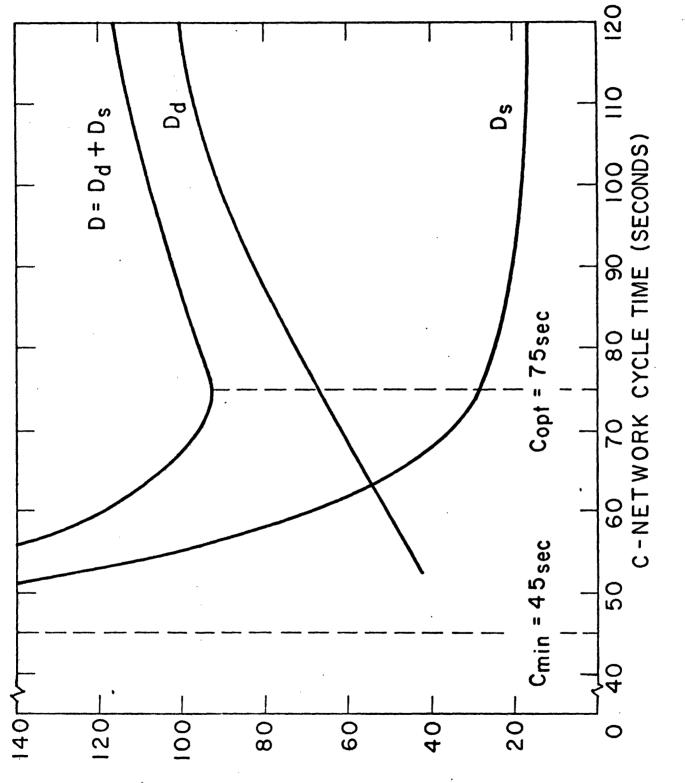


Figure 5: A. Platoon profile; B. Link performance function.

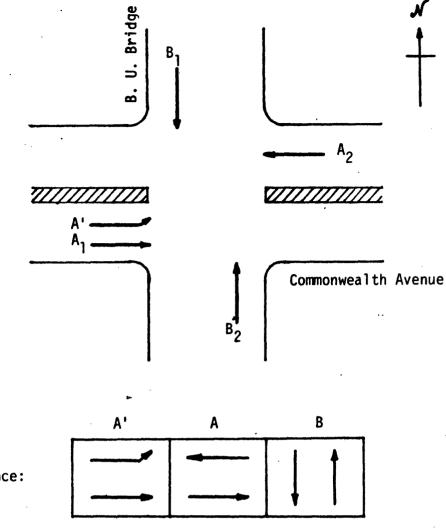
:1

ŧ



NETWORK RATES OF DELAY (veh x sec/sec)

Figure 6: Rate of total delay vs. cycle time in a synchronized network.



Phase Sequence:

4

C = 100 sec; L = 9 sec

Approach	q (vph)	s (vph)	g (cycles)	x	d (sec)	D (veh x sec/sec
A	240	1500	0.18	0.88	63.0	4.2
۹	1500	3000	0.54	0.925	30.7	12.8
A ₂	1200	3000	0.36	0.75	28.8	9.6
Bl	1050	3000	0.38	0.92	42.0	12.2
B ₂	500	3000	0.38	0.44	20.4	2.8
					TOTAL:	41.6

Figure 7A: Commonwealth / Boston University (3 phases).

ł

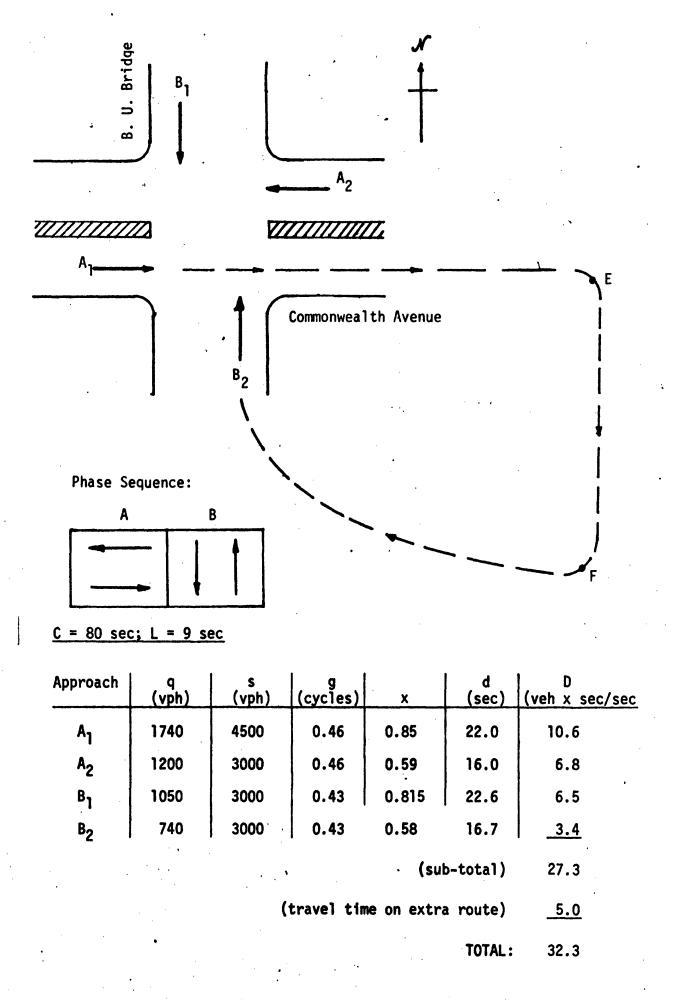


Figure 7B: Commonwealth / Boston University (2 phases).