

JB 80-01
TR 80-3

Paper Entitled
"Some Results on Computer Control of Urban Traffic"

From the Proceedings of
the IFAC Symposium on Large Scale Systems:
Theory and Applications

pp. 449-458

Toulouse, France
June 1980
Pergamon Press

SOME RESULTS ON COMPUTER CONTROL OF URBAN TRAFFIC⁺

J. S. Baras* and W. S. Levine*

*Department of Electrical Engineering, University of Maryland,
College Park, Maryland, USA*

Abstract. The problem of controlling vehicle traffic on an urban street network is a particular example of a large scale systems problem. In a reasonably typical network there will be one thousand or more traffic signals, there may be several hundred sensors (vehicle detectors) and the network is typically distributed over several hundred square miles. It is now quite common to control these systems primarily from a central computer.

The traffic control problem is modelled by a kind of "store and forward" network in which vehicles are "stored" in queues at the signals and "forwarded" in platoons. The data from detectors is modelled as a point process that is statistically related to the queues and platoons.

It would be very tempting to design optimal feedback control algorithms for these systems. However, the size of the problems makes this impossible. Instead, sub-optimal controls are developed. These sub-optimal controls use the minimum error variance estimator of the queues at each intersection as determined from the above model to improve the control provided by several intuitively derived sub-optimal controls. The sub-optimal controls have been tested using a good simulation of urban traffic. The sub-optimal controls give substantially better performance than benchmark open loop controls. Conditions under which the network can be decomposed into subnetworks without significant degradation in performance are also given.

Keywords. Traffic Control, Large Scale Systems, Computer Control, Modelling, Stochastic Control.

INTRODUCTION

The primary purpose of the traffic signal at an intersection is to prevent accidents by allocating the intersection to competing streams of traffic at different times. In urban areas, where the street network typically involves many nearby signalized intersections, it is obvious that some sort of coordination among signals is necessary to prevent the signals becoming a major impediment to the flow of traffic through the network. The purpose of this paper is to first describe a mathematical model of traffic flow on an urban network. The model is then used to derive signal control and coordination algorithms. These algorithms are traffic responsive in the sense that the signal transitions are determined by the signals from vehicle detectors located in the

streets. Finally, although hierarchical in form, the algorithms can be implemented in a collection of microprocessors, each with responsibility for a small subset of the total network.

To understand the motivation for this research it is necessary to briefly review the current state-of-the-art of urban traffic control. Nowadays, it is quite common in cities throughout the world for the coordination among traffic signals to be achieved by means of a central digital computer which controls the signals. With few exceptions, the coordination, or timing pattern, is based entirely on the time of day and on historical data. In most of the exceptions, the data from vehicle detectors is only used by the computer to determine which previously computed timing pattern to use.

Attempts to construct more directly traffic responsive network coordination algorithms have, until recently, been unsuccessful. That is, when the traffic responsive control was compared with a good open loop scheme applied to the same network, the open loop

⁺This research has been supported in part by the U.S. D.O.T. under contract DOT-OS-60134 and in part by the University of Maryland Computer Science Center.

*The authors are with the Department of Electrical Engineering, University of Maryland, College Park, Maryland 20742.

scheme, produced better or equivalent performance (Tarnoff, 1979), (Holroyd, 1972). Although it is certainly true that there are problems associated with vehicle detectors, the major problem has been the development of effective traffic responsive control algorithms. The model developed in this paper, besides leading to what are believed to be effective traffic responsive control algorithms, gives a plausible explanation for the previous failures.

Recently, two successful traffic responsive network traffic controls have been reported. One, called SCOOT (Robertson, 1979) has been tested in Glasgow and reportedly produced significant improvement over an optimized open loop control. Since SCOOT is proprietary very little is known about its operation. The other, called SCAT (for Sydney Coordinated Adaptive Traffic) reportedly produced 39% and 35% improvement over optimized open loop controls during the morning and evening peaks, respectively (Sims, 1979). SCAT will be discussed at somewhat greater length below since it has had a strong influence on the algorithms described in this paper.

The purpose of the research reported herein was to (a) provide a theoretical framework in which the above mentioned experience made sense and (b) to develop effective traffic responsive control algorithms. A mathematical model which reasonably achieves the first goal is described in the next section of this paper. The third section describes several algorithms for control at an isolated intersection and shows that these algorithms give better performance than an optimal open loop control. The fourth section describes a traffic responsive algorithm for a one way arterial and demonstrates that this algorithm gives better performance than an optimized open loop algorithm. It is also shown how this algorithm lends itself to decomposition of the arterial into sub-arterials. Then, the extension of this algorithm to networks is given. Finally, some conclusions and suggestions for further research are given.

Mathematical Model

Many mathematical models have been developed to describe the flow of traffic on an urban network. The best known and most widely used are the UTCS-1 Network Simulation Model (Lieberman and colleagues, 1977) now called NETSIM and the traffic flow models incorporated in TRANSYT (Robertson, 1969) and SIGOP (Anon., 1968). NETSIM is, however, too complex to use for the analytical determination of traffic responsive controls. SIGOP and TRANSYT are widely used to determine optimal open-loop controls and are very good at this. But, they do not include vehicle detectors. Thus, the first step in this research was to develop a mathematical model for vehicle flows on a network that would capture the statistical relationship between vehicle flows and detector signals and be reasonably tractable.

The model was partly determined by the initial conjecture that a traffic responsive control system needed to respond to detector data with a minimal delay. This meant that long time averages of detector data could not be used. Since the detector data is a sequence of 0's and 1's (1 if a vehicle is over the detector at the sampling time and 0 otherwise) this meant that the detector data had to be treated as a point process (Snyder, 1975).

The other major determinant of the model was that it is most common to treat urban traffic networks as a form of "store and forward" network. The basic notion is that traffic is "stored" in the form of queues just upstream from the traffic signals and is "forwarded" in the form of platoons on the links. Both TRANSYT and SIGOP, for example, are based on this idea. Thus, the construction of the model was begun by first developing mathematical models for the relation between queues and platoons and detector data. Since these models applied primarily on isolated links or intersections it was then necessary to model the interrelation of these links and intersections.

Thus, the network modelling problem was broken into three components:

1. Model the queue on each arm of an intersection.
2. Model the passage of each platoon over each detector.
3. Model the coordination of the above sub-models so as to determine overall network behavior.

Queue Model

The queue model is described in great detail in (Baras, Levine and Lin, 1979) and in more detail in (Baras, Levine and colleagues, 1977). Thus, the model is only summarized here.

The simplest practical queueing situation is the intersection of two one-way single-lane streets as illustrated in Fig. 1. Assume for simplicity that the signal operates on a known red-green cycle (no amber), and that there is only one detector located N vehicle lengths from the stop line. The observed signal from the detector will be denoted by $n^a(t)$.

$$n^a(t) = \begin{cases} 1 & \text{if a vehicle is over the detector} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

In practice, time is discretized with a small enough discretization interval (1/32 second in the UTCS in Washington) for each vehicle to be over the detector for several samples. For simplicity, it is assumed here that the data are sampled so that each vehicle produces exactly one pulse.

Let

$$\lambda(k,t) = P_r [n^a(t) = 1 \text{ given that there are } k \text{ vehicles in the queue and the time is } t] \quad k=0,1,2,\dots,N \quad (2)$$

$$z(t) = \text{the number of vehicles in the queue at time } t \quad (3)$$

Assume that

$$\lambda(k,t) = \lambda(t) = \lambda_r u_u(t-\tau) + \lambda_g (1-u_u(t-\tau)); \quad (4)$$

$$k = 0,1,2,\dots,N-1$$

$$\lambda(N,t) = 0 \quad (5)$$

$$\text{where } u_u(t) = \begin{cases} 1 & \text{if upstream traffic signal is red} \\ 0 & \text{if upstream traffic signal is green} \end{cases}$$

and τ is the known average time for a vehicle to get from the upstream stop line to the detector.

It is important to understand the implications of the above assumptions. First, when the queue contains N vehicles, no more vehicles can cross the detector which leads to Eq. (5). Of course, in very heavy traffic flow conditions the queue may very well exceed N by a substantial amount. Although the model could be extended to handle this case, this will not be done here. This limits the validity of this model to less than very heavy flow conditions. It is believed that the moderate to light flow conditions provide the most opportunity for large improvement due to traffic responsive control because in these flow conditions one sees the largest amount of randomness. Heavy flow conditions are relatively deterministic and, therefore, relatively predictable and, therefore, relatively open-loop controllable.

Secondly, it is assumed that the time required for the lead vehicle in the queue to proceed to the next detector is very predictable (always τ). This is a common and reasonable assumption in moderate to light flow. Finally, the two level approximation to the arrival rate λ given in Eq. (4) corresponds quite closely to the similar assumption in SIGOP. It would be reasonably easy to improve the approximation in Eq. (4) by including the effects of traffic signals further upstream. The result would be very similar to the model in TRANSYT.

To complete the queuing model, imagine there is a detector at the stop line producing an unobserved departure point process $n^d(t)$.

$$n^d(t) = \begin{cases} 1 & \text{if a vehicle departs at time } t \\ 0 & \text{otherwise} \end{cases}$$

Let

$$\mu(k,t) = P_r [1 \text{ vehicle departs given } k \text{ vehicles in queue at time } t]$$

$$\mu(k,t) = \begin{cases} \mu(1-u_i(t)) & ; k>0 \\ 0 & k=0 \end{cases}$$

$$\text{where } u_i(t) = \begin{cases} 1 & \text{if traffic signal at this intersection is red in through direction} \\ 0 & \text{if traffic signal at this intersection is green in through direction} \end{cases} \quad (6)$$

This actually completes the model for queuing at one arm of the intersection. To see this, note that if the switch times of the traffic signals are known

$$P_r [z(t+1)=j | z(t)=i] = \begin{cases} 0 & j < i-1, j > i+1 \\ \lambda(i,t)\mu(i,t) + (1-\lambda(i,t))(1-\mu(i,t)), & i=j \\ \lambda(i,t)(1-\mu(i,t)), & j=i+1 \\ \mu(i,t)(1-\lambda(i,t)), & j=i-1 \end{cases} \quad (7)$$

$$P_r [n^a(t)=1 | z(t)=i] = \begin{cases} \lambda(t) & i < N \\ 0 & i = N \end{cases} \quad (8)$$

The model can be placed in a more suggestive form by defining

$$x_k(t) = \begin{cases} 1 & \text{if there are } k \text{ vehicles in the queue at time } t \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

where $k = 0, 1, 2, \dots, N$

It is then straightforward (Baras, Levine and Lin, 1979a) to show that Eqs. (7) and (8) are equivalent to

$$\begin{aligned} \underline{x}(t+1) &= \underline{Q}^T(t)\underline{x}(t) + \underline{w}(t) \\ n^a(t) &= \underline{\lambda}^T(t)\underline{x}(t) + v(t) \end{aligned} \quad (10)$$

$$\text{where } \underline{\lambda}^T(t) = [\lambda(t) \lambda(t) \dots \lambda(t) 0]$$

$$Q_{ij}^T(t) = P_r [z(t+1)=j | z(t)=i] \text{ (see Eq. (7))}$$

and $\underline{w}(t)$ and $v(t)$ are "noise" processes. More precisely $\underline{w}(t)$ and $v(t)$ are martingale difference sequences with respect to the σ -algebra generated by the sequences

$$\{n^a(0), n^a(1), \dots, n^a(t-1)\} \text{ and}$$

$$\{\underline{x}(0), \underline{x}(1), \dots, \underline{x}(t)\}.$$

It should be apparent from Eqs. (7) and (8) that the model is a controlled Markov chain with partially observable state. It is straightforward to extend the model to complicated intersections (including two way intersections in which left turning vehicles complicate the departure rate μ), to intersections with multiple detectors and to utilize velocity data from detectors. Some of this is reported in Baras, Levine and Lin (1979a) and in Baras, Levine and colleagues (1977). It should also be noted that the model depends on only three parameters and is relatively insensitive to the values of λ_r and λ_g . The model is sensitive to the value of μ . The parameter μ is an important parameter in most traffic flow models but is usually called the saturation flow by traffic engineers. Methods to

adaptively estimate μ are currently under investigation. They are based on the use of the platoon model described below.

Platoon Model

The platoon passage model can be described in terms of the simple situation illustrated in Fig. 1. Suppose that, some short time previously, the signal just upstream of the figure has turned green. This releases the queue at the signal which then flows, as a platoon, over the detector in Fig. 1. It is assumed that the arrival of the lead vehicle in the platoon is easily predictable and that the fundamental problem is to estimate the passage of the last vehicle in the platoon. While the platoon is crossing the detector, it is assumed that time headways between successive vehicles (time between successive l 's of $n^a(t)$) satisfy a lognormal distribution

$$P_f(h) = \begin{cases} \frac{1}{\sigma h \sqrt{2\pi}} \exp \frac{(\ln h - \alpha)^2}{2\sigma^2}, & h \geq 0 \\ 0 & , h < 0 \end{cases}$$

Once the platoon has past, it is assumed that traffic is free-flowing and satisfies a displaced exponential distribution.

$$P_{nf}(h) = \begin{cases} \beta \exp(-(h-\tau)\beta), & h \geq \tau \\ 0 & , < \tau \end{cases}$$

It is also assumed that successive headways are independent. Thus, headway statistics at the detector are completely described by the probability density of headway

$$p(h) = \psi(t) p_f(h) + (1-\psi(t)) p_{nf}(h)$$

where $\psi(t)$ denotes the switch from following to non-following headways and is determined by the upstream traffic signal.

Although this has been done very briefly, this virtually completes the formulation of the platoon passage model. The piece of the platoon flow model that is not given above is the initialization of the model which is determined by the upstream queue estimate. The model is completely described in Baras, Dorsey and Levine (1979b) as well as Baras, Levine and colleagues (1977). It should be noted that the model is closely related to various other models of platoon flow (see references cited for details).

Coordination

Although the queueing model described above applies only to a single arm of a single intersection, it already contains the basis for combining these single intersection models into a network model. Clearly, the queue model is coupled to its nearest upstream neighbor by means of $\lambda(t)$ and, therefore, $\lambda(t)$ contains the network coordination (see

Eq. (4)). Thus, assume that

- (1) the network consists entirely of one way streets with signals at every intersection,
- (2) there is a detector N_i vehicle lengths from the stopline on every link, $i = 1, 2, \dots, I$,
- (3) λ_{ri} and λ_{gi} are known for every link $i = 1, 2, \dots, I$.

Then, once the traffic signal values, $u_j(t)$, $j = 1, 2, \dots, J$ ($u_j(t) = 0$ if signal is green and 1 if it is red) are known the network coordination is completely specified by the $\lambda_i(t)$ $i = 1, 2, \dots, I$ for every link with

$$\lambda_i(t) = \lambda_{ri} u_{ji}(t - \tau_i) + \lambda_{gi} [1 - u_{ji}(t - \tau_i)] \quad i = 1, 2, \dots, I \quad (11)$$

where $u_{ji}(t) = \begin{cases} 1 & \text{if traffic signal at inlet to link } i \text{ is red} \\ 0 & \text{if traffic signal at inlet to link } i \text{ is green.} \end{cases}$

τ_i = average time required for a vehicle to get from the inlet to link i to the detector on link i

λ_{ri} and λ_{gi} are related to the departure rates at the upstream intersection and the percentage of turning traffic.

Several observations are in order concerning this model of network coordination. First, this model of network coordination corresponds approximately to the models used in SIGOP and TRANSYT. Consider a more detailed comparison to TRANSYT. TRANSYT is based on the following assumptions:

- (1) All traffic signals are periodic with the same period (called the cycle), T .
- (2) The arrival flow at the inlets to the network (see Fig. 2) has exponentially distributed inter-arrival times with constant λ . That is, $\lambda_i(t) = \lambda_i$ for all inlet links.

Each cycle is then divided into fifty sub-intervals. Then, if the traffic signals are specified ($u_j(t)$ specified $j = 1, 2, \dots, J$, all t) it is straightforward to calculate $\lambda_i(t)$ for every link $i = 1, 2, \dots, I$. The result will be a picture similar to Fig. 3 for a typical $\lambda_i(t)$. In fact, TRANSYT

includes several other more sophisticated features including a platoon dispersion model. The essential point to be made here is that Eq. (11) can be viewed as a fairly simple approximation to the above described aspects of TRANSYT and that the λ_{ri} and λ_{gi} should be computed as in TRANSYT. Thus, one sensible view of the proposed model is that it is an approximation to TRANSYT or SIGOP with a very detailed queueing model superimposed at each intersection.

Secondly, notice that conservation of vehicles is only satisfied on the average (the λ 's satisfy conservation of vehicles if computed as described above but they are only averages). This is reasonable in urban networks where parking causes vehicles to enter and exit the system or equivalently to vanish and reappear at random. Similarly, the correlation between adjacent queues has been ignored in the above model since $\lambda_i(t)$ contains no information about the upstream queue. This could be fixed, at the expense of greater complexity, in several ways. The easiest is probably to use a TRANSYT approximation to $\lambda_i(t)$ (fifty values per cycle). Better, but much more complex, would be to let

$$\lambda_i(k_r, k_g, t) = p_r \mu_r(k_r, t - \tau_i) + (1 - p_g) \mu_g(k_g, t - \tau_i) \quad (12)$$

where k_r, k_g = number of vehicles in through queue and perpendicular queue respectively at time t
 p_r, p_g = probability a vehicle does not turn from through and perpendicular queues respectively
 τ_i = as before.

Clearly, this is hopelessly complicated and so is abandoned. A third option would be to utilize the platoon passage model developed above to model the network coordination. This would also be quite complicated and so one is left with the relatively simple model in Eq. (11).

Third, the above model appears to capture the essential difficulty in designing traffic responsive controls for grid networks. That is, the "downstream" demands and, therefore, the "downstream" controls are highly correlated to the "upstream" signal. However, this correlation occurs with a delay corresponding to the possibly large travel time between the two intersections. For example, in Fig. 2, the controls at node 8 and node 1 will be dominant factors in determining the optimal control at node 2. Thus, calculation of the optimal control at node 8 requires accurate prediction of the demand that will occur at node 2 a total of at least $(\tau_6 + \tau_7)$ time units into the future. To further complicate matters, in a true grid network, there is flow in both directions so "upstream" and "downstream" are ambiguous designations. To take an extreme example, in Fig. 2, node 8 is "downstream" from node 2 via the path traversing links 15-10-18-19-3-22. The point is that the tight, but delayed, coupling between intersection controls creates a difficult control problem. Furthermore, the more complex and accurate models discussed above intensify this coupling.

Fourth, the model suggests an obvious method for decomposing a large network into sub-networks. Any links for which $\lambda_i(t) = \lambda$, a constant, can be cut without loss of information. Such a link can be regarded as an output link for one subnetwork and an input

link for another subnetwork. This rule of thumb is in fact used by the developers of TRANSYT to decompose traffic networks (private communication from P. Hunt of TRRL, 1979).

Fifth, and finally, the initial assumptions can now be seen to be not restrictive. Two way streets can be modelled as two one-way streets. Intersections that do not have signals simply change λ_{ri} and λ_{gi} in Eq. (11). Missing detectors result in the queue on that link being dropped from the model. Note that this is important because detector reliability is something of a problem. And, finally, λ_{ri} and λ_{gi} can be computed from knowledge of impact flows to the network and saturation flows in the network, data that is also used by TRANSYT or SIGOP.

Although the equations will not be written out explicitly, the model developed above is clearly just a controlled Markov chain with partially observable states. Of course, the chain has a huge number of states so it is impractical to attempt to compute an optimal control. The following sections of this paper develop sensible sub-optimal controls.

CONTROL OF ISOLATED INTERSECTIONS

There are two reasons to begin the discussion of controls for networks by examining the degenerate case of a single isolated intersection. First, the structure and solution of the isolated intersection problem is a valuable guide to the solution of the network problem. In fact, the solution given here for the network problem is an extension of the solution for the isolated intersection. Second, in practice there are a significant number of intersections that can be, and are, controlled in isolation.

Thus, consider the isolated intersection of two one way streets. Assume each arm of the intersection contains a detector located as in Fig. 1. Let $z_\ell(t)$ ($\ell = 1, 2$) denote the number of vehicles in the queue on arm ℓ of the intersection at time t and let $n_\ell^a(t)$ denote the corresponding detector signal. The model contained in Eqs. (7) and (8) then applies to each arm of the intersection independently provided that the control signal $u(t)$ (see Eq. (6)) is known. Unfortunately, if a traffic responsive, or equivalently, a feedback control is used the queues and detector signals on the two arms of the intersection become statistically dependent.

This statistical dependence means that, in fact, it is impossible to separate the problems of estimating the state (the size of both queues) and determining the controls. That is, the separation principle does not apply. However, because of the inherent advantages in separation, the separation of state estimation and control will be imposed subsequently by limiting the class of

admissible controls. Once this separation is imposed, one can write the dynamics of the intersection as

$$\underline{x}_\ell(t+1) = Q_\ell^T(t, u(t)) \underline{x}_\ell(t) + \underline{w}_\ell(t) \quad \ell = 1, 2 \quad (13)$$

$$n_\ell^a(t) = \lambda_\ell^T(t) \underline{x}_\ell(t) + v_\ell(t)$$

Although many measures of performance are used in urban traffic control, the aggregate delay is almost always considered important. In practice, the traffic signal is not allowed to switch too often because (1) there is usually a minimum allowable green time for each arm of the intersection and (2) each transition introduces some additional delay. Thus, it is reasonable to attempt to minimize the performance criterion below

$$J = E\{\sum_{t=1}^T C_1^T \underline{x}_1(t) + C_2^T \underline{x}_2(t) + C_3 (\Delta u(t))^2\} \quad (14)$$

$$\text{where } \Delta u(t) = \begin{cases} +1 & \text{when the signal facing arm 1} \\ & \text{changes from red to green} \\ -1 & \text{when the signal facing arm 1} \\ & \text{changes from green to red.} \end{cases}$$

and C_1 and C_2 are weights which allow one direction to be given priority over the other while C_3 penalizes too many switches.

Finally, the set of admissible controls will be limited to those control laws which depend only on the optimal estimate of the state based on the model in Eq. (13) and the observations $n_\ell^a(t)$, $\ell = 1, 2$. The minimum error variance estimate of $\underline{x}_\ell(t)$ given

$n_\ell^a(0), n_\ell^a(1), \dots, n_\ell^a(t-1)$, which will be denoted by $\hat{\underline{x}}_\ell(t|t-1)$, has been derived in

(Baras, Levine and Lin, 1979a) and in (Baras, Levine and colleagues, 1977). The derivations, explicit algorithms which can be realized in a microprocessor and test results are given in the above references and so, will not be given here. Here, it is sufficient to indicate that

$$\hat{\underline{x}}_\ell(t+1|t) = \underline{f}_\ell(t, \hat{\underline{x}}_\ell(t|t-1), n_\ell^a(t), u(t)) \quad (15)$$

where \underline{f}_ℓ is a known function given

explicitly in the above cited references. Thus, Eq. (15) replaces Eq. (13) as the mathematical description of the dynamics for the control problem. And, the admissible controls become

$$\Delta u(t) = g(\hat{\underline{x}}_1(t|t-1), \hat{\underline{x}}_2(t|t-1), t) \quad (16)$$

where $g(\cdot, \cdot, \cdot)$ is an arbitrary function.

Replacing $\underline{x}_\ell(t)$ by $\hat{\underline{x}}_\ell(t|t-1)$ ($\ell = 1, 2$) in Eq. (14) results in a well defined stochastic optimal control problem with dynamics given by Eq. (15), control by

Eq. (16) and performance given by Eq. (14). This problem can be solved, in principle, by dynamic programming but such a solution could not be implemented in practice. Thus, although research is continuing on this problem, at present it is not feasible to use the truly optimal control even for the isolated intersection, much less for the network.

Instead, several sub-optimal controls were developed. The basic idea behind the development of these sub-optimal controls was to use the above-mentioned $\hat{\underline{x}}_\ell(t|t-1)$ as the queue estimates in control algorithms that had been previously developed by traffic engineers. Two such control algorithms were used. The first is a fairly standard control algorithm that was previously analyzed approximately by Darroch, Newell and Morris (1964). The algorithm is described by assuming that the light has just turned from red to green on arm one. The light is kept green for long enough to empty the current queue and then remains green as long as the interarrival time between subsequent vehicles does not exceed a value β_ℓ . In practice, there is a maximum allowable green time as well. The second is an improved version of this algorithm developed by Michalopoulos and Stephanopoulos (1979). Both obviously require effective queue estimates.

The performance of both of these suboptimal control algorithms was evaluated by means of the UTCS-1 network simulation (Lieberman and colleagues, 1977). A single intersection was simulated. A standard open-loop control procedure, due to Webster (1958), was used as a benchmark. The results are summarized in tables 1 and 2 below.

TABLE 1 Results of Single Intersection Control Tests for Heavy Traffic on Both Arms of the Intersection

	Average Speed (mi/hr)	Average Delay Per Vehicle (secs)	Stops Per Vehicle
benchmark	14	42.4	.84
M. + S.	15.5	36.6	.65
D.N. + M.	17	32	.67

TABLE 2 Results of Single Intersection Control Tests for Moderate Traffic on One Arm and Light Traffic on the Other

	Average Speed (mi/hr)	Average Delay Per Vehicle (secs)	Stops Per Vehicle
benchmark	24.6	15.3	.43
M. + S.	29.3	10.3	.13
D.N. + M.	30.3	9.3	.06

These tests will be reported in more detail in Baras, Levine and colleagues (1979c). However, the salient point here is that, using estimators based on the model described in this paper, these control algorithms made significant improvements over an excellent open loop control. This provides validation for the model as well as suggesting effective sub-optimal controls for the network case.

CONTROL OF ONE WAY ARTERIALS

An arterial is simply an urban street, complete with traffic signals, that carries a relatively large volume of through traffic. The interesting case is when the cross streets intersecting such an arterial carry random traffic. For example, consider the network of Fig. 2 and imagine that traffic on all the vertical links satisfies the condition $\lambda_i(t) = \lambda_i$. Then, for purposes of determining traffic signal timings, the network can be decomposed into three independent subnetworks. Each of these subnetworks (the arterial consisting of links 1,2,3 and 4 with their associated cross streets, for example) can be regarded as a one way arterial with random cross traffic.

Such one way arterials occur quite frequently in practice. Even two way arterials frequently have much larger flow in one direction and so can be treated as though they were one way. Good open loop control schemes for these one way arterials are known. The closed loop control problem is considerably simpler than it is for a more general network. The principle reason for this is that the well defined direction of flow allows control decisions to be made sequentially in both time and space.

Since it is clearly impossible to compute the optimal control based on the model developed earlier in this paper the main effort has been development of effective suboptimal controls. The suboptimal control that was developed is based on three ideas:

- (1) Control the first upstream node of the arterial as though it is an isolated intersection,
- (2) Control the nominal coordination of the downstream nodes in an open loop fashion based entirely on the average time to traverse the links,
- (3) Adjust the timing of the downstream nodes, within limits, according to the local estimates of queues at the node.

To see the control scheme more clearly, consider Fig. 4. Node 1 is the upstream node and this intersection is controlled, as though it is an isolated intersection, by either of the two algorithms described in the previous section. Suppose the transition from red to green occurs on the light facing link one at $t = t_0$. Suppose link two has length ℓ_2 and mean free speed v_2 . Then, the nominal switch time from red to green of the signal at node two facing link two will be

$(t_0 + \ell_2/v_2)$. In fact, this switch time will be advanced slightly in practice to take care of vehicles moving slightly faster than the average and vehicles queued on link two. This nominal "offset" is quite good on the average. However, it does not account for variations in the queues on links two and nine. This variation can be accurately estimated by means of the queue estimator described earlier. Based on this estimate, the nominal offset ℓ_2/v_2 is adjusted by up to $\pm \sigma$ seconds. The opposite transition, from green to red facing the arterial, is determined by the isolated intersection algorithm modified to insure that the next red to green transition can occur at the proper time. Obviously, the same basic scheme, with proper choice of offset, is used for the intersections further downstream.

The above algorithm was tested by means of the UTCS-1 simulation (Lieberman and colleagues, 1977) on an arterial corresponding to Fig. 4. As a benchmark, the well known open loop procedure of Little, Martin and Morgan (1964) was used. The results are shown in Table 3 below.

TABLE 3 Tests of Arterial Control Algorithm for Moderate Traffic on the Arterial and Light Crossing Traffic

	Average Speed (mi/hr)	Stops/Vehicle	Average Delay Per Vehicle (secs)
benchmark	17.4	.67	20.6
proposed control	18.9	.62	16.2

The tests summarized above involved fairly small amounts of turning traffic. It is clear that large amounts of turning traffic will cause substantial degradation in the performance of the above algorithm. However, if large amounts of turning traffic exist, the arterial can be broken into sub-networks each of which is controlled independently. This idea can be explained more clearly by considering the example tested. Consider the arterial illustrated in Fig. 4 and the simulation that resulted in Table 4. Change the simulation so that about 30% of all traffic at each intersection turns. This causes the volume of traffic on the arterial to decrease as it moves downstream. Three control algorithms were tested for this case. First, the benchmark algorithm was used. Second, the proposed algorithm was used to control the full length of the arterial. Third, the arterial was decomposed into two sub-networks by cutting (for control purposes only) link four. Nodes 1, 2 and 3 were controlled via the proposed algorithm with node 1 as the upstream node. Nodes 4 and 5 were controlled by the same

basic algorithm independently from the rest of the network. The results of the test are given in Table 4.

TABLE 4 Tests of Arterial Control
Algorithm for Significant Amounts
of Traffic Turning Off the Arterial

	Average Speed (mi/hr)	Stops Per Vehicle	Average Delay Per Vehicle (secs)
benchmark	17.3	.58	17.4
proposed control one network	18.5	.56	14.1
proposed control two sub- networks	18.8	.45	13.4

Several comments are in order. First, the algorithm proposed here, complete with decomposing the network into subnetworks whenever the average volume on adjacent arterial links is substantially different, is quite similar in spirit if not detail to SCAT (Sims, 1979). The algorithm clearly performs quite well in the tests described here. In similar tests in which volume on the arterial increased as the distance downstream increased (due to large amounts of traffic turning into the arterial) the benchmark open loop algorithm gave the best performance.

The difficulty in making the proposed algorithm work well when the largest demand occurs at the downstream node illustrates the problem of traffic responsive network control extremely well. The volume of traffic at an intersection basically determines the nominal period (cycle) of the traffic signal at that intersection. The heavier the volume the longer the period. Any pair of intersections that are to be coordinated must have the same nominal cycle. Thus, only one node in a coordinated arterial can be allowed to operate as an isolated intersection. That one node should be the one with heaviest volume and will be called the leader node. When that node is downstream from other nodes in the arterial, these upstream nodes must predict the signal transition of the leader node so they can compute their offsets. The required predictions can be quite far into the future, on the order of several cycles for a long arterial.

Based on the model described in this paper the optimal, in the minimum error variance sense, predictors of the signal transitions of the leader node are known. And, the optimal prediction for any time greater than one cycle depends only on the average volumes, the known nominal offsets and the traffic signal settings. Thus, the detector data is not useful for prediction times longer than one cycle and so one might as well use open loop prediction. Of course, one might object that the model described here is over simplified

and ignores the coupling between queues and "nearby" detector data. However, the estimators based on this simplified model have been tested and work quite well. The implication is that a more elaborate model will not produce significant improvement in prediction accuracy.

This suggests that the best traffic responsive control procedure when the leader node on an arterial is downstream would be the following:

- (1) Determine the nominal cycle and split (percentage of green allocated to arterial) at the leader node by Webster's Method (1958), the benchmark for the isolated intersection.
- (2) The nominal offsets are determined, as before, from the mean free speed and length of the links.
- (3) The nominal signal transitions are adjusted, based on the optimal queue estimates, by an amount no greater than $\pm \sigma$ seconds. This local adjustment is carried out in isolation from the rest of the network.

The above control algorithm is a form of open loop local feedback optimal (OLLFO) control scheme and will be referred to as the OLLFO control. It will be discussed in greater detail in the next section.

CONTROL OF GRID NETWORKS

The general grid network presents a much more difficult coordination problem than the arterial because there is no unambiguous definition of "upstream" and "downstream". However, the OLLFO scheme proposed in the previous section is certainly applicable to the general network. In fact, a related idea was tried as part of the UTCS experiment in Washington, D.C. (Tarnoff, 1975). The idea was called, Critical Intersection Control (C.I.C.) and was basically to (a) use an essentially open loop control for the network as a whole and (b) to make "small" adjustments based on local detector data at a few "critical" intersections. The conclusion drawn from the tests was that C.I.C. did not produce significant improvement.

There are two major differences between C.I.C. and the OLLFO algorithm proposed here. First, the local feedback control algorithm proposed here is based on much better estimates of the queue at the local intersection. Second, the local feedback control would be applied at most, if not all, intersections. If this is not done one would expect much of the gain from C.I.C. at one intersection to be nullified at nearby intersections.

The local feedback optimal control proposed here is based on solving a slightly modified version of the isolated intersection problem defined by Eqs. (14), (15) and (16). However, the performance criterion would be

$$J_{LFO} = E \left\{ \int_{t=t_0-\sigma}^{t_0+\sigma} C_2^T \hat{x}_1(t|t-1) + C_2^T \hat{x}_2(t|t-1) \right\} \quad (17)$$

where t_0 is the nominal switching time

$\pm\sigma$ is the maximum allowable change from the nominal switching time.

For reasonable values of σ , say ± 5 seconds, this optimal control problem, or a close approximation, should be solvable.

Clearly, the best that one can expect from such limited local feedback control is that one or two more vehicles per lane can be squeezed through the intersection without delay. Even in heavy traffic this would represent a potential 5-10% improvement over open loop control. Since the real advantage to feedback control is believed to occur in moderate and light traffic this is believed to be a significant potential improvement.

CONCLUSIONS

The proposed network control algorithm has not been tested on a grid network. The tests of a version of the algorithm were, however, successful on a simulated arterial. Thus, the proposed algorithm is believed to be practically promising. Similarly, the proposed model for traffic flow on an urban network seems to capture many of the special features of urban traffic flow that have been observed in practice. Thus, the model is believed to have potential use in the theoretical investigation of algorithms for traffic control.

At present, urban traffic control systems are believed to be a fruitful subject for research. The typical urban network is clearly a large scale system problem with a rich structure. There is some evidence that the practice, in the form of SCAT and SCOOT, may now be leading the theory. Thus, additional theoretical work is indicated. Our group is currently investigating a number of theoretical aspects of the problem including (a) the establishment of good performance estimates for the OLLFO algorithm, (b) methods for adaptively improving the estimates of λ_{ri} and λ_{gi} on the network and (c) the establishment of good bounds on the degradation in performance due to decomposition of a network into subnetworks.

REFERENCES

- Anon. (1968). SIGOP Traffic Signal Optimization Program Users Manual. Peat, Marwick, Livingston and Co., New York.
- Baras, J.S., W.S. Levine and T.L. Lin (1979a). Discrete time point processes in urban traffic queue estimation. IEEE Trans. on AC, Vol. AC-24, pp. 12-27.
- Baras, J.S., W.S. Levine, A.J. Dorsey and T.L. Lin (1977). Advanced Filtering and Prediction Software for Urban Traffic Control Systems. Final Report, U.S. Dept. of Transportation, Contract # DOT-OS-60134.
- Baras, J.S., A.J. Dorsey and W.S. Levine (1979b). Estimation of traffic platoon structure from headway statistics. IEEE Trans. on AC, Vol. AC-24, pp. 553-559.
- Baras, J.S., W.S. Levine, A.J. Dorsey and T.L. Lin (1979c). Improved Software for Computerized Urban Traffic Control Systems. Dept. of Transportation, Contract DOT-OS-60134, Final Report, (to appear).
- Darroch, J.N., G.F. Newell and R.W.J. Morris (1964). Queues for a vehicle actuated traffic light. Operations Research, V. 12, pp. 882-895.
- Holroyd, J. (1972). Comparison Between a Dynamic Signal Plan Generation System of Area Traffic Control and a Fixed-Time System. TRRL Report LR 487, Transport and Road Research Laboratory, Crowthorne, Berkshire, United Kingdom.
- Lieberman, E., D. Wicks and J. Woo (1977). Network Flow Simulation for Urban Traffic Control System-Phase II, Vol. 4 User's Manual for UTCS-1 Network Simulation Model. Federal Highway Administration, Washington, D.C.
- Little, J.D.C., B.V. Martin and J.T. Morgan, (1964). Synchronizing traffic signals for maximal bandwidth. Operations Research, Vol. 12, pp. 864-912.
- Michalopoulos, P.G. and G. Stephanopoulos, (1979). An algorithm for real time control of critical intersections. Traffic Engrg. and Control, January 1979, pp. 9-15.
- Robertson, D.I. (1969). TRANSYT: A Traffic Network Study Tool. Road Research Laboratory Report LR 253, Crowthorne, Berkshire, United Kingdom.
- Robertson, D.I. (1979). Traffic models and optimum strategies of control—a review. Proc. of the International Symposium on Traffic Control Systems, Vol. 1-Resource Papers, publ. by Institute of Trans. Studies, Univ. of Calif. at Berkeley, pp. 262-288.
- Sims, A.G. (1979). The Sydney coordinated adaptive traffic system. Proc. of the Engrg. Found. Conf. on Research Directions in Computer Control of Urban Traffic Systems, publ. by Am. Soc. of Civil Engrs., pp. 12-27.
- Snyder, D.L. (1975). Random Point Processes. John Wiley and Sons, New York.
- Tarnoff, P.J. (1975). The results of FHWA urban traffic control research: an interim report. Traffic Engrg., Vol. 45, pp. 27-35.
- Tarnoff, P.J. (1979). Concepts and strategies urban street systems. Proc. of the Int'l. Symposium on Traffic Control Systems, Vol. 1-Resource Papers, publ. by Inst. of Trans. Studies, Univ. of Calif. at Berkeley, pp. 1-12.
- Webster, F.V. (1958). Traffic signal settings. Road Research Paper No. 39, British Road Research Laboratory.

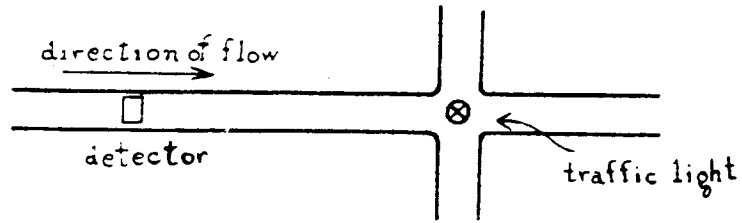


Fig. 1 Detector location on one arm of an intersection.

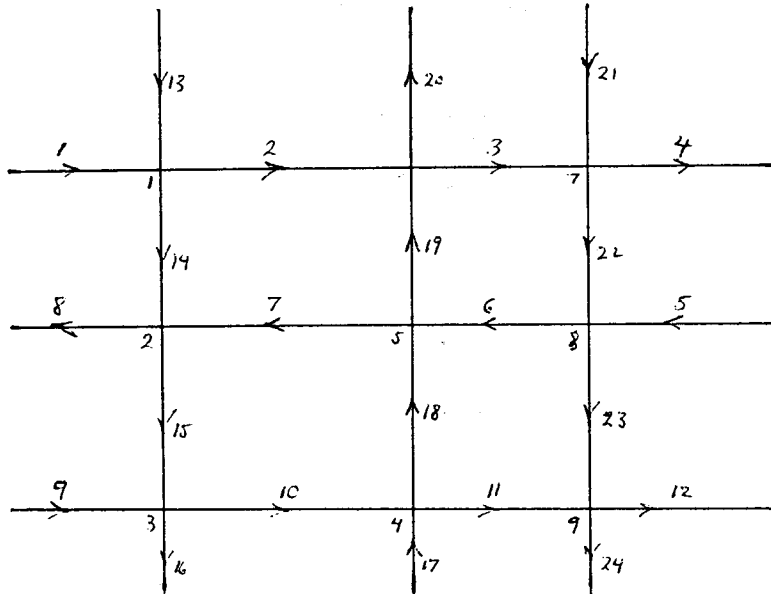


Fig. 2 Example of a one-way grid network. Links 1, 9, 13, 21, 5, 17 are inlet links. Links 8, 16, 24, 12, 4, 20 are outlet links. $I = 24$, $J = 9$.

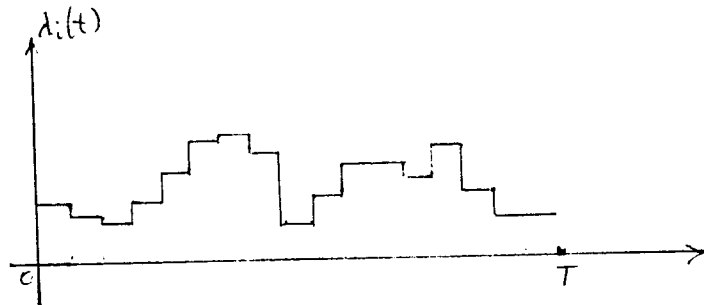


Fig. 3 Typical plot of $\lambda_i(t)$ vs. t for one cycle.

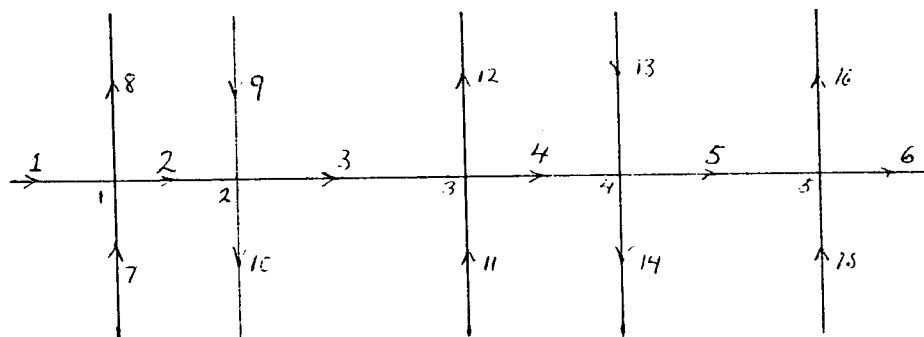


Fig. 4 Example of a one-way arterial.