erally results in improvements in all impacts, minimizing passenger delays is one of the two highest ranked objective functions.

2. Priority-vehicle delay: Minimizing priority-vehicle delay resulted in allocating all available green time to the study arterial. Large penalties were thus accrued to vehicles approaching on a cross street. More thorough tests should be performed by using this objective before further guidelines are recommended. Tests on arterials that have intersecting bus routes or are in a network should prove illuminating.

3. Stops: Because the TRANSYT model computes stops from uniform delays only, the use of this objective alone may result in oversaturation of some intersections; thus, its use as a single-impact objective is not recommended. Stops may, however, be used in combination with delay as in the earlier TRANSYT 6 model—this gives a better balanced objective function.

4. Fuel consumption: Using fuel consumption as a single-impact objective generally minimizes fuel consumption for the study section. Stops are nearly commensurably decreased because the amount of fuel consumed is heavily dependent on the number of stops. Both total delay and priority-vehicle delay are decreased, but not as much as in the total-delay-based optimization. Because of its overall favorable impacts, minimizing fuel consumption is the other choice for highest ranking single-impact objective.

5. Vehicle emissions: Most of the tests indicated very small changes in vehicle emissions. Part of the problem may be due to the fact that, for the test section, most vehicle emissions result from cruise, not delay. In general, this objective function is more closely related to vehicle flows than to passenger flows. Therefore, impact savings for priority vehicles are usually sacrificed for better overall vehicle performance. This objective function may be helpful in decreasing the vehicle-emission impacts of strategies in operating environments in which vehicle emissions are a severe problem.

6. Extensions: The tests of alternative single-impact objective functions are a first step in a more thorough analysis of the potentially powerful multiple-objective function. Combinations of impacts can be used, and different weights can be assigned to each. An example of this would be to assign equivalent monetary costs to each impact and then set signals to minimize a net cost.

ACKNOWLEDGMENT

This research was supported by the Office of University Research, U.S. Department of Transportation.

REFERENCES

1. D. I. Robertson. TRANSYT: A Traffic Network Study Tool. U.K. Road Research Laboratory, Crowthorne, England, Rept. LR 253, 1969.
2. A. D. May. Some Fixed-Time Signal-Control Computer Programs. Proc., Traffic Control Symposium, Monte Carlo, Monaco, Sept. 1974.
3. P. J. Claffey. Running Cost of Motor Vehicles as Affected by Road Design and Traffic. NCHRP, Rept. 111, 1971.
4. P. Kunselman and others. Automobile Exhaust Emission Modal Analysis Model. U.S. Environmental Protection Agency, Rept. 460/3-74-005, 1974.
5. K. Train. TDFP Demand Model. Travel-Demand Forecasting Project, Institute of Transportation Studies, Univ. of California, Berkeley, Internal Rept., 1976.
6. P. P. Jovanis, A. J. Kruger, and A. D. May. Analysis of Traffic Management Impacts and Demand Responses on Surface Streets and Freeways. Institute of Transportation Studies, Univ. of California, Berkeley, 1977.

Publication of this paper sponsored by Committee on Urban System Operations.

Optimization of Large Traffic Systems

Stanley B. Gershwin, Electronics Systems Laboratory, Massachusetts Institute of Technology
Paul Ross and Nathan Gartner, Office of Research, Federal Highway Administration
John D. C. Little, Operations Research Center, Massachusetts Institute of Technology

A program for optimizing most of the significant traffic engineering quantities while allowing for the fact that drivers can and do change their routes and modes of travel is proposed. The program requires the person origin-destination table as input and gives the best traffic-signal settings, reserved-lane assignments, and ramp-metering rates as outputs. The optimization can be for minimum person time, for minimum fuel consumption, or for a combination of both. The optimization was coded by using arbitrary but reasonable models for the components and run for a hypothetical network. Sample results are given. The example problems described required from 2.5 min (system optimizing) to 3 min (user optimizing) central processing unit time per complete iteration.

All of the signal-optimization methods that are now available use fixed traffic demands—i.e., they assume that changes in the signal settings will not produce significant changes in the actual traffic. This assumption is made not only in those methods that optimize isolated signals, e.g., Webster's method (1), but also in the computer methods that optimize entire networks of signals simultaneously, such as TRANSYT (2, 3, 4), SIGOP (5), and MITROP (6, 7, 8). However, it is clear that signal settings do have an effect on traffic assignments (9, 10).

The research described in this paper is an attempt to examine the feasibility of carrying traffic optimization one or two steps beyond this level and to produce a technique that will allow the traffic engineer to determine the correct signal settings, freeway ramp-metering rates, and reserved-lane assignments while taking into account the fact that automobile drivers will change their routes and passengers will change their modes of travel. Travelers can also change their trip frequencies and destinations in response to traffic conditions and traffic engineers can make other changes to the system (such as intersection improvements), but we have not included such factors in the optimization model. The procedure we describe here is similar to others (11, 12); however, we provide numerous options and we believe our traffic equilibrium technique may be more accurate. Other authors have combined demand studies with assignment (13, 14, 15) and with modal split (16, 17, 18).

This research is an extension of that reported elsewhere (19). It is only a feasibility study, not a finished and polished program; the intent was to determine whether such large-scale, comprehensive optimizations are possible, not to actually produce a consumer product. Because of this emphasis on the development of a pilot program, there are a few caveats that should be remembered when examining the numerical results:

1. The component submodels for traffic assignment, signal optimization, and modal split have not been calibrated or validated. The models that were used are reasonable and have been fairly well justified in the literature but they are not meant to be authoritative; research is continuing in all these areas.
2. The computer code used for the numerical examples is not in a polished and finished form. The coding was done to examine the feasibility of the concept; it does not operate as smoothly and conveniently as a finished program would and may even contain minor bugs.

OVERALL OPTIMIZATION PLAN

The overall optimization plan is a fairly simple iteration among three separate subprograms—traffic assignment, signal optimization, and modal split. These three subprograms can use adaptations of established and well-known computer packages [such as UROAD (20) for the traffic assignment and TRANSYT for the signal optimization]. However, because of the iterative nature of the optimization, it is important that these routines execute very quickly; therefore, they should be written especially for this use. In the experiments carried out here, specially written subprograms were used, except that the MITROP signal-optimization program was also used in some runs.

Traffic Assignment

The most crucial element of the optimization is the traffic-assignment program. In addition to the usual feature of estimating the average traffic on all links of a network from fixed origin-destination (O-D) tables, the traffic assignment must be able to make these assignments on the basis of two different criteria—system optimization and user optimization. System-optimized traffic means that all traffic takes the route that is best for the system as a whole. That is, some drivers go out of their way and take longer or more expensive (or both) routes to avoid congesting certain streets so that other drivers will have much shorter or cheaper (or both) trips. In this way, the total time spent by all travelers in the system is a minimum, but such altruism is rare among real automobile drivers. Nevertheless, it is important that the traffic-assignment components have the ability to make such assignments for reasons that will be explained below.

User-optimized traffic assignments are the products of programs that assume that each individual driver does as well as he or she individually can; i.e., any other route that can be chosen takes at least as long as the one actually chosen. It follows that all routes that carry any traffic between a given O-D pair have the same trip times and the routes that are not used take even longer. There is reason to believe that actual traffic approximately distributes itself according to this user-optimization principle. The principles of user and system optimization, in the context of traffic, are described by Wardrop (21).

Signal Optimization

This traffic-assignment package can be combined with a signal-optimization package to optimize the signal settings in a traffic network while allowing for the ability of traffic to change its route distribution. In Figure 1, the traffic-assignment program shown in box 2 iterates back and forth with the signal optimization in box 3 until the traffic flows cease to change appreciably from iteration to iteration and the stopping criterion (box 4) is satisfied. The signal-optimization program should accept the average traffic volume on the streets as input and produce the best signal settings as output. In addition, it should provide output that allows the traffic-assignment program (box 2) to reflect the effects of signal synchronization and coordination. Of the available signal-network optimization packages, only MITROP has this ability.

Modal Split

Travelers will change their mode of travel if they perceive a benefit in doing so. In stochastic systems such as traffic, this perception takes a relatively long time but modal shift is nevertheless a real phenomenon. Changes in the mode of travel can be included in this optimization by adding another interation loop, as shown in Figure 2. Box 2 represents the entire program of Figure 1; it requires only an O-D table (broken down by mode of travel) and it gives, as output, the correct traffic flows, signal settings, and travel times for each O-D pair by mode. (We assume that the traffic-assignment subprogram can recognize several different kinds of traffic—e.g., automobiles, car pools, and buses—and evaluate the total travel time associated with trips by each mode.) Box 3, the modal-split determination, uses these travel times for each mode to estimate what fraction of persons will travel by each mode for each O-D pair. Any suitable program will serve for the modal-split module (box 3) if it can take the information provided by box 2 (travel times by mode for each O-D pair) and produce a reasonable estimate of the fractions of persons choosing each mode.

This three-way iteration converges to an optimal solution, but we are unable to prove whether this optimum is global. In the test cases, convergence was quite rapid at first, coming to within 1 or 2 percent of the eventual modal split in two or three iterations, but after that was relatively slow.

Florian (22) has solved the modal-split problem (for two modes, automobile and public transit, only) by a different iterative procedure. He guesses the transit travel times and solves an elastic demand problem in the automobile sector. This leads to new transit travel

Figure 1. Flow chart for calculation of signal settings and assignments.

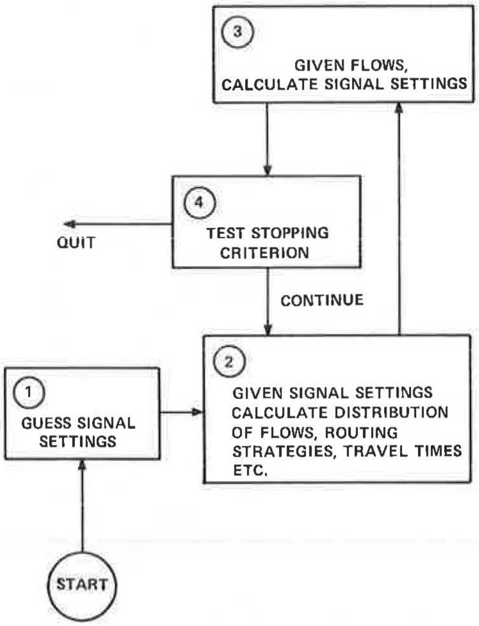
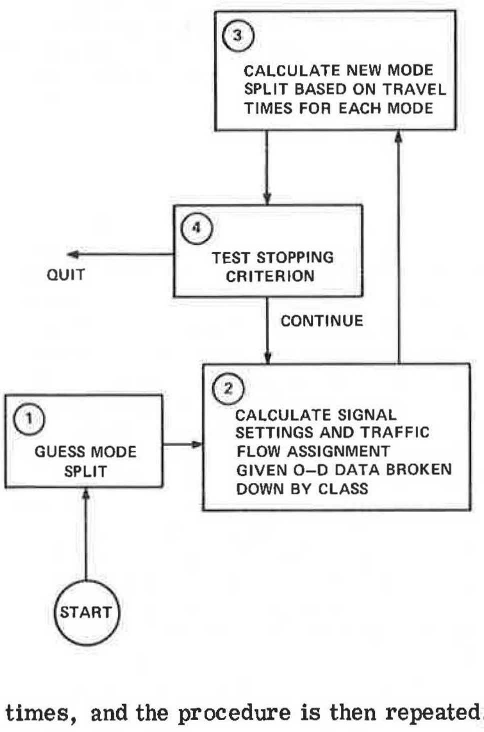


Figure 2. Flow chart for calculation of modal split.



times, and the procedure is then repeated.

Reserved Lanes

If, for example, we run the optimization of Figure 1 by using the traffic assignment in the system-optimizing mode, the program might assign only buses to a certain street and require automobiles and car pools to take longer routes to their destinations. This might be the only way to achieve the minimum person delay in the system. Of course, in real, user-optimizing traffic, ordinary automobiles would crowd onto this street and degrade the overall performance of the system unless the street was reserved exclusively for buses.

Lane reservations are made by an extension of this principle. The network geometry must be entered into the computer such that each lane that is a candidate for reserved status is coded as a separate link and suitable nodes are located so that traffic can enter and leave at reasonable locations. A link-and-node representation of a hypothetical freeway corridor is shown in Figure 3. Link 36, which joins nodes 15 and 29, is a (possibly) reserved lane. Traffic entering the freeway at node 16 cannot cross the regular lanes (link 37) until further downstream at node 29. Computer runs are made by using the traffic assignment in both the user- and system-optimizing modes, and the lanes where a significant change in the mix of traffic occurs are reserved for the traffic component that dominates in the system-optimized mode. (This is not a true optimization procedure; it only indicates that the option should be evaluated.) This comparison could be automated, but there are so many unquantifiable value judgments involved that we have chosen to leave the lane-reservation process as a manual judgment based on the program output and local knowledge.

Ramp Metering

The determination of optimal ramp-metering rates requires a compound optimization procedure of the type described by Gartner and others (23).

DETAILS OF ACTUAL COMPONENTS

To test computationally the concepts outlined above, we have constructed actual computer programs. Because this research was in the nature of a pilot study, no great effort was made to achieve ultimate accuracy and perfection in the programming but all components were at least reasonable. The parameters used were, in some cases, only reasonable guesses.

Traffic Assignment

The traffic-assignment subprogram is the heart of the optimization; it contains the traffic model and an optimization subroutine.

The traffic model must give the average travel time on each link while allowing for the presence of several different types of vehicles and several different kinds of streets (e.g., arterials, freeways, and freeway ramps). If the traffic model gives these travel times accurately and simply, the optimization portion will have little difficulty in assigning traffic to the various streets so that the total objective function is a minimum (system optimum) or the trip times by alternative paths are equal (user optimum). Because this traffic-assignment subprogram is used many times during the course of the overall program, it is important that the optimization routine operate with exceptional speed.

The traffic model itself has two components: the constraints and the objective function. The constraints describe the relations among the variables of the system. In the model developed for this study, the constraints were

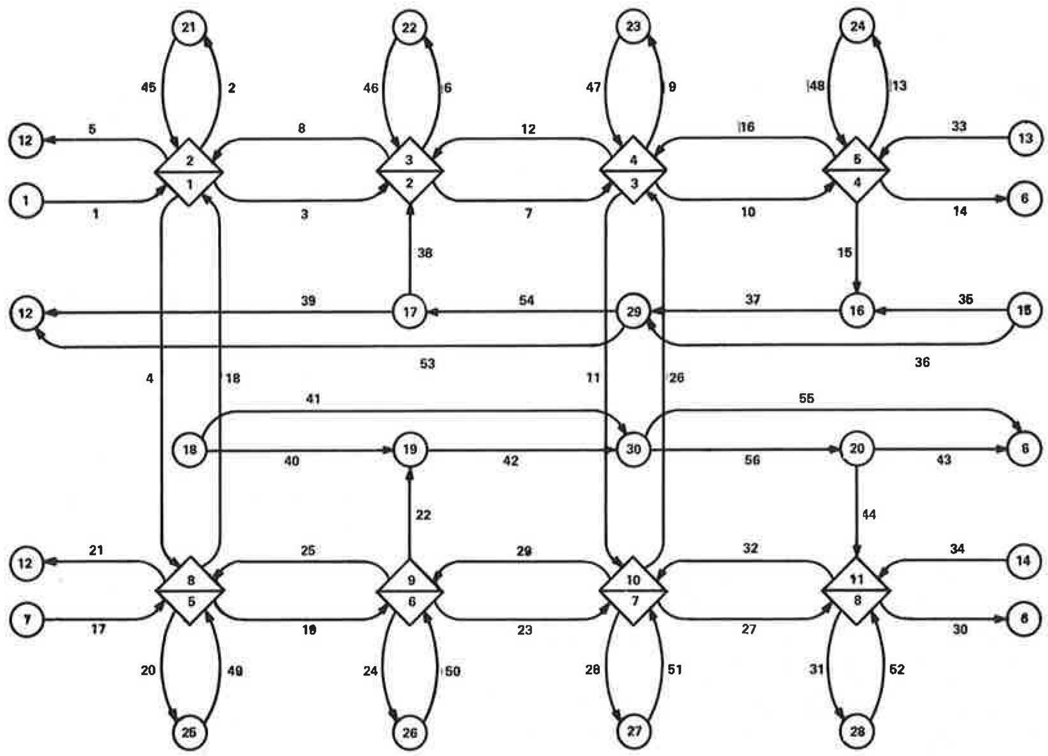
1. Nonnegative flows

$$\phi_{ij}^{(n)} \geq 0 \tag{1}$$

for all i, j, n;

2. Net outflow from source node k equals O-D volume

Figure 3. Network that has freeway lanes separated.



$$\sum_{\substack{\text{links } \beta \\ \text{leaving} \\ \text{node k}}} \phi_{\beta j}{}^{(n)} - \sum_{\substack{\text{links } \alpha \\ \text{entering} \\ \text{node k}}} \phi_{\alpha j}{}^{(n)} = r_{kj}{}^{(n)} \quad (j \neq k); \tag{2}$$

3. Net inflow to destination node k equals the sum of attractions

$$\sum_{\substack{\text{links } \alpha \\ \text{entering} \\ \text{node k}}} \phi_{\alpha k}{}^{(n)} = \sum_{m \neq k} r_{mk}{}^{(n)} \tag{3}$$

and, at transfer node k, net outflow equals net inflow

$$\sum_{\substack{\text{links } \beta \\ \text{leaving} \\ \text{node k}}} \phi_{\beta j}{}^{(n)} - \sum_{\substack{\text{links } \alpha \\ \text{entering} \\ \text{node k}}} \phi_{\alpha j}{}^{(n)} = 0 \quad (j \neq k); \tag{3a}$$

4. Total passenger flow (P_i) defined on link i

$$P_i = \sum_j [\phi_{ij}{}^{(1)} + 2.5\, \phi_{ij}{}^{(2)} + 25\, \phi_{ij}{}^{(3)}]; \tag{4}$$

and

5. Total passenger-automobile-equivalent volume defined on link i

$$\phi_i = \sum_j [\phi_{ij}{}^{(1)} + \phi_{ij}{}^{(2)} + 3\phi_{ij}{}^{(3)}] \tag{5}$$

where

$\phi_{ij}{}^{(1)}$ = passenger-automobile flow (vehicles/h) on link i destined for node j,
$\phi_{ij}{}^{(2)}$ = car-pool flow (vehicles/h),
$\phi_{ij}{}^{(3)}$ = bus flow (vehicles/h), and
$r_{kj}{}^{(n)}$ = number of vehicles per hour of mode n originating at node k and destined for node j.

In constraint 4, we have assumed that private passenger automobiles carry one occupant, car pools average 2.5 occupants, and buses average 25 passengers. In constraint 5, we have assumed that buses have the same effect on traffic as three ordinary automobiles or car pools.

Objective Function: Time

The objective function (T) is the sum of the person travel times on all links

$$T = \sum_i P_i \tau_i \tag{6}$$

where P_i = total number of persons per hour using link i as defined by constraint 4 and τ_i = average travel time on link i. Different formulas were used for these travel times depending on whether link i represented a freeway, a freeway entry ramp, or an arterial street.

On freeway links, the average travel time depended only on the congestion on the link

$$\tau = l[1/v_0 + (0.3k_{max} - 1/v_0)(\phi/s)^6] \tag{7}$$

where

l = length of link,
k_{max} = jam density,
s = capacity, and
v_o = free speed on link.

On freeway entrance ramps, there is one delay at the metering signal and another at the point of merge with the freeway. A reasonable model for the metering delay can be derived from the theory for (M/D/1) queues as

$$[2 - (\phi/M)]/2M[1 - (\phi/M)] \tag{8}$$

where M = metering rate and ϕ = link flow (24). (This metering delay was not included in the computer program whose results are described below.)

Merge delay occurs on the ramp when an automobile waits for an acceptable gap in the freeway flow. It is reasonable to assume that the automobiles on the ramp wait for a gap of length D; thus, we can calculate the average merge delay from the queuing theory for (M/M/1) queues as

$$\{\phi'/[\exp(\rho'D) - 1] - \phi\}^{-1} \tag{9}$$

where ϕ' = freeway curb-lane flow that inhibited the merge process and ρ' = vehicular density in that lane. If we combine transit time, metering delay, and merge delay, the average travel time on a freeway entrance ramp is

$$\tau = t_0 + \{[2 - (\phi/M)]/2M[1 - (\phi/M)]\} + \{[\phi'/\exp(\rho'D) - 1] - \phi\}^{-1} \tag{10}$$

On arterials, there are significant delays at signals. The results of the optimization will be most accurate if the arterial-delay function includes, in detail, the effects of cycle length, split, and offset, but this accuracy will be at the expense of a considerable number of iterations between the traffic-assignment package and the signal-optimization package. At the other extreme, if the arterial-delay function ignores these variables, only one iteration will occur in each step and the overall program will execute with considerably increased speed. Although it is not absolutely necessary, logical consistency suggests that this arterial-delay function in the traffic-assignment module should be the same as that used in the signal-optimization module.

We used two different levels of detail in the arterial-delay function. The simpler arterial-delay function was taken from Webster:

$$\tau = t_0 + 0.45\left(\{c(1 - g)^2/[1 - (\phi/s)]\} + [\phi/gs(gs - \phi)]\right) \tag{11a}$$

where c = signal cycle length and g = fraction of the cycle that the link experiences a green display. The use of this function implies local optimization only, i.e., at each signal independently.

The more detailed arterial-delay function was chosen to be compatible with MITROP. (The MITROP program uses piecewise linear approximations to this curve.) This function takes account of the network effects of signal synchronization and coordination.

$$\tau = t_0 + \{\gamma^2/2p[1 - (\phi c/ps)]\} + \phi/(gs - \phi) \times \{1.25[(gs - \phi)/g^2 s^2 c]^{\frac{1}{3}} + 2.25[(\phi/gs) \times (gs - \phi)/g^2 s^2 c] + 0.008(\phi/gs)^3\} \tag{11b}$$

where γ = platoon arrival time at the downstream signal and is a function of the offset on the link and p = platoon length on the link. Both of these values are available from the MITROP signal-network-optimization program; this function could not be used if the values of γ and p were not produced by the signal-optimization program.

Objective Function: Fuel Consumption

We included fuel consumption in the system optimal objective function; vehicle emissions can be included in a similar manner. The total objective function was written in the form

$$W_t T + W_f F \tag{12}$$

where

W_t and W_f = user-specified parameters,
T = total travel time in the system as defined in Equation 6, and
F = total fuel consumed in the system.

The total fuel is the sum of the fuel consumed on the different links where, again, different formulas must be used to evaluate the fuel consumed on different kinds of links.

On freeways, the fuel consumed on link i (ξ_f) was taken to be

$$\xi_f = l[\phi^{(1)}G^{(1)}(v) + \phi^{(2)}G^{(2)}(v) + \phi^{(3)}G^{(3)}(v)] \tag{13}$$

where $G^{(n)}(v)$ = fuel consumption (L/km) of vehicles of mode n for the average speed v on the link. The fuel-consumption functions have been investigated by Claffey (25). We have used polynomial approximations to his data:

$$G^{(1)} = G^{(2)} = 0.285\,468 - (9.852\,41 \times 10^{-3})v + (1.864\,59 \times 10^{-4})v^2 - (1.465\,63 \times 10^{-6})v^3 + (4.413\,77 \times 10^{-9})v^4 \tag{14a}$$

$$G^{(3)} = 0.304\,520 - (1.229\,78 \times 10^{-2})v + (3.010\,67 \times 10^{-4})v^2 - (2.736\,72 \times 10^{-6})v^3 + (9.507\,19 \times 10^{-9})v^4 \tag{14b}$$

Because Claffey did not publish results for buses, Equation 14b was fitted to his results for two-axle, six-tired trucks.

On freeway entrance ramps, the principal effects are decelerating to a stop and then accelerating back to speed plus the idling time at the metering signal and the merge point. The total fuel consumed on entry ramps (ξ_{er}) was taken to be

$$\xi_{er} = \sum_n \phi^{(n)}\left[H^{(n)}(v) + I^{(n)}\{[2 - (\phi/M)]/2M[1 - (\phi/M)]\} + \left(I^{(n)}\{\phi'/[\exp(\phi' t^*) - 1]\} - \phi\right)^{-1}\right] \tag{15}$$

where the fuel consumed during the deceleration-acceleration cycle was

$$H^{(1)} = H^{(2)} = -1.827\,92 \times 10^{-2} + (1.862\,66 \times 10^{-3})v - (3.646\,46 \times 10^{-5})v^2 + (3.910\,79 \times 10^{-7})v^3 - (1.433\,24 \times 10^{-9})v^4 \tag{16a}$$

and

$$H^{(3)} = -4.008\,24 \times 10^{-5} + (3.433\,29 \times 10^{-4})v + (1.068\,39 \times 10^{-5})v^2 + (7.827\,85 \times 10^{-8})v^3 - (1.766\,52 \times 10^{-9})v^4 \tag{16b}$$

where $H^{(n)}(v)$ = fuel consumption (L/km) of vehicles of mode n for the average speed v during the deceleration-acceleration cycle and the fuel consumed during idling was

$$I^{(1)} = I^{(2)} = 2.19 \text{ L/h} \tag{17a}$$

and

$$I^{(3)} = 2.46 \text{ L/h} \tag{17b}$$

On signalized arterials, we took the fuel consumption (ξ_a) to be

Figure 4. Person O-D table for example problems.

	TO 6	12	21	22	23	24	25	26	28
FROM 1	870								225
7	1005					135			
13		420		240			510		
14		579	112					135	
15		5250					135	135	
18	5850								
21	537		---						225
22				---			135		
23					---				180
24		315				---			
25	372				135		---		
26	315					90		---	
27				135					
28					90				---

$$\xi_a = (k_1 l + k_2 \tau)\,[\phi^{(1)} + \phi^{(2)} + 1.523\phi^{(3)}] \tag{18}$$

where k_1 = 0.1020 L/vehicle-km and k_2 = 0.000976 L/vehicle-s. This is based on the results of Evans and Herman (26). The 1.523 factor for buses is $G^{(3)}$ (48 km/h)/$G^{(1)}$(48 km/h).

Optimization Portion of Traffic Assignment

We used the Cantor-Gerla algorithm (27) for the optimization portion of the test traffic-assignment routine; other traffic-assignment optimization routines, such as those based on the Frank-Wolfe method or the convex-simplex method, could also be used. [These methods have been reviewed by Gartner (28).] An experiment that used a novel extension of user optimization to nonseparable cost functions had promising results. The master step in the system-optimization algorithm can be thought of as finding a solution to a set of nonlinear equations and inequalities—the Kuhn-Tucker conditions (29). Suppose instead, we satisfy a different set of nonlinear equations and inequalities: We find flows to equate travel times on paths defined by the extremals already generated. If this procedure converges, it creates flows that equate travel times on paths used between origins and destinations—the user-optimization solution.

Signal-Optimization Package

We incorporated the MITROP optimization program into some of the runs; all components operated well, and good signal progressions were established. However, most of the computational work used a very simple signal-optimization procedure. Green time was assigned to each signal approach in proportion to the degree of saturation on that approach. This ignores the effects of cycle length, signal offset, and platoon formation but, nevertheless, provides a fast and easy solution; it is adequate when the signals are far enough apart so that platoons dissipate in traveling between them (30, 31). It is, of course, inadequate where the effects of signal synchronization and coordination are important. Because, in this method, the concepts of offset and platoon length are not used, values for the γ and p variables required in Equation 11b are not available; only the Webster delay formula (Equation 11a) could be used. All the results reported in this paper use the Webster delay function; cycle lengths were fixed at 60 s.

Modal Split

There does not appear to be a consensus as to what factors affect the choice of mode of travel. In the absence of such a consensus, it is proper to choose the simplest reasonable way of representing modal choice. We chose the well-known logit model (22, 32)

$$r_{ij}^{(n)} = [R_{ij}/w^{(n)}]\exp[-\alpha^{(n)} - \beta^{(n)}T_{ij}^{(n)}] \Big/ \sum_n \exp[-\alpha^{(n)} - \beta^{(n)}T_{ij}^{(n)}] \tag{19}$$

where

Mode	α	β
Automobile	0.00	0.010
Car pool	0.15	0.011
Bus	0.25	0.013

R_{ij} = total number of persons that will travel from origin i to destination j by all modes (the person O-D table),

$w^{(n)}$ = average occupancies of the modes (taken as 1, 2.5, and 25 in these examples and in constraint 4), and

$T_{ij}^{(n)}$ = trip time (seconds from origin i to destination j) via mode n.

The α and β parameters are arbitrary choices that seemed reasonable.

This closed form for the modal-choice function is attractive because of its computational efficiency.

RESULTS OF SAMPLE RUNS

To demonstrate the feasibility of the concept of optimization of traffic systems, the test programs described above were run by using the network shown in Figure 3 and the person O-D data shown in Figure 4. Both the network and the O-D table are hypothetical constructs chosen to illustrate typical conditions.

Optimization in the user-optimizing mode in which persons are free to change mode of travel approximates the way real traffic would occur in the network. If the hourly traffic volumes are as shown, 553.36 person-h would have been spent in the network and 2323 L of gasoline would have been consumed each hour.

The same O-D table was also run in the fuel system-optimizing mode with free modal choice. That is, each person chose his or her mode of travel by its apparent utility to himself or herself but took the best route to conserve fuel. Although real people do not choose their routes this way, it provided a sort of limit on the possible fuel consumption and served as the basis from which reserved lanes were determined. In this run, 2155 L of fuel were consumed; no combination of traffic improvements alone would be likely to reduce long-term fuel consumption below this value. To achieve lower fuel consumption, either coercive travel restrictions or changes in the average fuel consumption of vehicles or changes in the public perception of the utility of different modes would be required.

Examination of the traffic volumes in this system-optimal assignment showed that buses were not being assigned to the outer two lanes of some freeway links and that automobiles and car pools were restricted from the inner lane of the freeways. One freeway entrance ramp was being used only by buses, and the other was not being used at all. Consequently, a third run was made in which drivers were allowed to choose their own routes (user optimum), but there were extensive restrictions on what types of vehicles could use which links. Only buses were allowed on the innermost freeway lane (links 36, 41, 53, and 55). Entrance ramp 22 was closed to all traffic, and entrance ramp 15 was closed to all but buses. Buses were prohibited from the outer lanes on freeway links 35 and 39 because they were expected to use the inner lane (links 36 and 53). Modal splits were fixed at the values that resulted from the run that represented the situation that would be expected just after these lane restrictions were introduced and before travelers had an opportunity to change modes. Quite dramatic fuel savings resulted from the new traffic restrictions; even some travel time would have been saved as indicated below (1 L = 0.264 gal).

Run Conditions	Person-Time in Network (person-h/h)	Total Fuel Consumed (L/h)
Present value (user optimum, no reserved lanes, modal shifts allowed)	553.36	2323
Best possible fuel consumption; no changes in network (system optimal, fuel basis, modal shift allowed)	507.33	2155
Expected after reserved lanes introduced (user optimal, no modal shift)	530.42	2084
Expected eventually (user optimal, modal shift allowed)	507.69	2151

Travelers can be expected to continue to use their old mode of travel only until they perceive an advantage in changing. We have assumed that, after lane restrictions are instituted, it will take only a few days for the traffic to establish new routes but several months for the travelers to arrive at a new modal-choice equilibrium. Because the mode choices are based on personal utility, they will not necessarily improve overall fuel use. In fact, when we allowed travelers to choose their new modes of travel in the last run, the total fuel consumption returned to the optimal value from the second run. (The difference between these values is within the limits of the convergence tests; i.e., they are essentially equal.) In general, we would expect the new fuel consumption with lane restrictions to be above the optimum value; the lane restrictions used here turned out to be unexpectedly efficacious.

The differences between the third and fourth runs appear to be counterintuitive. When the reserved bus lanes are first introduced (the third run), in effect, a large part of the corridor capacity is lost; the most pronounced effect is that automobiles on the freeway are forced to move at slow, fuel-efficient speeds.

Many travelers whose trips lie along the length of the corridor switch to buses, which reduces the freeway congestion somewhat in the fourth run. The freeway itself becomes less congested, and automobiles move at higher, less efficient speeds; however, the shift to buses produces a small net saving of fuel on the freeway itself, although elsewhere the network remains congested and both automobiles and buses moving across the corridor are delayed. When both automobiles and buses are delayed, there is a strong shift from buses to automobiles. The net effect for the whole network is a decrease of 8 buses/h; this increased use of automobiles for trips across the corridor more than cancels the fuel savings along the corridor.

It is not meaningful to compare the eventual fuel consumption with the artificially low fuel consumption of the third run. That consumption could be maintained only if travelers were enjoined from changing modes of travel. The correct comparison is between the first and fourth runs and shows that the introduction of bus lanes has saved not only fuel but also time.

CONCLUSIONS AND PROSPECTS

This project has shown that a partial traffic-engineering optimization of traffic systems is feasible. Such a computer program should be quite useful for indicating the overall effects of contemplated changes and for highlighting points where improvements would be most effective. We feel that this has been the first attempt to look at second-order effects in the traffic system.

Vehicle emissions could be evaluated by methods analogous to those outlined here for fuel consumption. The resulting program would evaluate the total person time, fuel consumed, and vehicle emissions produced in a network and indicate what can be done to reduce any one or any combination of these factors (27, 33).

The algorithms that were used in this study seem suitable for large-scale use, but careful attention to clean, error-free programming and the minimization of execution time and memory are required. The utility of the Cantor-Gerla algorithm and the overall iterative scheme has been shown, but there are both obvious and subtle ways to improve their efficiency. (For example, the first few iterations—those through the traffic assignment and the signal optimization—require only approximate solutions, but the programs, nevertheless, produce fully accurate results based on the current approximate input. Much quicker, less accurate functions could be used for these first few iterations.) The example problems described in this paper required from 2.5 min (system optimizing) to 3 min (user optimizing) central processing unit (CPU) time per complete iteration. (When there are lane restrictions, as used in the second and third runs, the network is considerably simpler, and a complete modal-split iteration required only 0.38 min of CPU time.) We believe that clean and efficient programming will reduce these running times by an order of magnitude, which will make it possible to analyze corridors that have approximately 150 intersections and 500 links within reasonable computer time and storage requirements (perhaps three complete iterations in 30 min of CPU time and 250 K of core storage on an IBM 360/65).

REFERENCES

1. F. V. Webster. Traffic Signal Settings. U.K. Road Research Laboratory, Crowthorne, England, Technical Paper 39, 1957.
2. D. I. Robertson. TRANSYT: A Traffic Network Study Tool. U. K. Road Research Laboratory, Crowthorne, England, Rept. LR 253, 1969.
3. D. I. Robertson. TRANSYT Method for Area Traffic Control. Traffic Engineering and Control, Vol. 2, No. 6, Oct. 1969.
4. F. A. Wagner. TRANSYT Method for Network Optimization. Proc., ITE, 42nd Annual Meeting, Sept. 26, 1972.
5. SIGOP: Traffic Signal Optimization Program—A Computer Program to Calculate Optimum Coordination in a Grid Network of Synchronized Traffic Signals. Traffic Research Corporation, New York, Sept. 1966; NTIS, Springfield, VA, PB 173 738.
6. N. H. Gartner, J. D. C. Little, and H. Gabbay. MITROP: A Computer Program for Simultaneous Optimization of Offsets, Splits, and Cycle Times. Traffic Engineering and Control, Aug.-Sept. 1976, pp. 355-359.
7. N. H. Gartner, J. D. C. Little, and H. Gabbay.

Optimization of Traffic Signal Settings by Mixed-Integer Linear Programming—Part 1: The Network Coordination Problem and Part 2: The Network Synchronization Problem. Transportation Science, Vol. 9, 1975, pp. 321-363.
8. 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.
9. M. J. Maher and R. Akcelik. The Redistributional Effects of an Area Traffic Control Policy. Traffic Engineering and Control, Sept. 1975.
10. N. H. Gartner. Area Traffic Control and Network Equilibrium. In Traffic Equilibrium Methods (M. A. Florian, ed.), Springer-Verlag, New York, 1976.
11. J. A. Charlesworth. The Calculation of Mutually Consistent Signal Settings and Traffic Assignment for a Signal-Controlled Road Network. Proc., 7th International Symposium on Transportation and Traffic Theory, Kyoto, Japan, July 1977, pp. 545-561.
12. R. E. Allsop and J. A. Charlesworth. Traffic in a Signal-Controlled Road Network: An Example of Different Signal Timings Inducing Different Routings. Traffic Engineering and Control, Vol. 18, May 1977, pp. 262-264.
13. M. Beckman, C. B. McGuire, and C. Winston. Studies in the Economics of Transportation, Yale Univ. Press, New Haven, CT, 1956.
14. M. Florian and S. Nguyen. A Method for Computing Network Equilibrium With Elastic Demands. Transportation Science, Vol. 8, No. 4, Nov. 1974.
15. M. Florian, S. Nguyen, and J. Ferland. On the Combined Distribution-Assignment of Traffic. Transportation Science, Vol. 9, No. 1, Feb. 1975.
16. M. R. Wigan and T. J. G. Bamford. An Equilibrium Model of Bus and Car Travel Over a Road Network. U.K. Transport and Road Research Laboratory, Crowthorne, England, Rept. LR 559, 1973.
17. J. C. Yu and R. K. Giguere. Optimizing Transportation Facility Arrangements. Transportation Engineering Journal, Nov. 1973.
18. D. Ghahrman, A. K. C. Wong, and T. Au. Perturbations of a Multimodal Network Model for Urban Transportation Planning. Trans., IEEE, Systems, Man, and Cybernetics, Vol. SMC-4, No. 3, May 1974.
19. S. B. Gershwin, J. D. C. Little, and N. H. Gartner. Computer Assisted Traffic Engineering Using Assignment, Optimal Signal Setting, and Modal Split. Research and Special Programs Administration, U.S. Department of Transportation, Rept. DOT-TSC-RSTA-78-10, 1978.
20. E. R. Ruiter. Network Equilibrium Capabilities for the UMTA Transportation Planning System. In Traffic Equilibrium Methods, Montreal, 1974, Springer-Verlag, Berlin, 1976, pp. 183-199.
21. J. G. Wardrop. Some Theoretical Aspects of Road Traffic Research. Proc., Institute of Civil Engineers, Pt. 2, 1952.
22. M. Florian. A Traffic Equilibrium Model of Travel by Car and Public Transit Modes. Transportation Science, Vol. 2, No. 2, May 1977, pp. 166-179.
23. N. H. Gartner, S. B. Gershwin, and J. D. C. Little. Supply Modeling for Urban Traffic Management. Paper presented at the International Symposium on Travel Supply Models, Univ. of Montreal, Nov. 1977.
24. H. A. Taha. Operations Research: An Introduction. Macmillan, New York, 1971.
25. P. J. Claffey. Running Costs of Motor Vehicles as Affected by Road Design and Traffic. NCHRP, Rept. 111, 1971.
26. L. Evans and R. Herman. A Simplified Approach to Calculations of Fuel Consumption in Urban Traffic Systems. Traffic Engineering and Control, Aug.-Sept. 1976.
27. D. G. Cantor and M. Gerla. Optimal Routing in a Packet-Switched Computer Network. Trans., IEEE, Computers, Vol. C-23, No. 10, Oct. 1974, pp. 1062-1069.
28. N. H. Gartner. Analysis and Control of Transportation Networks by Frank-Wolfe Decomposition. Proc., 7th International Symposium on Transportation and Traffic Theory, Kyoto, Japan, July 1977, pp. 591-623.
29. H. W. Kuhn and A. W. Tucker. Nonlinear Programming. Proc., 2nd International Berkeley Symposium on Mathematical Statistics and Probability (J. Neyman, ed.), Univ. of California Press, Berkeley, 1951.
30. F. V. Webster and J. A. Hillier. Traffic Control Systems. In Traffic Engineering Practice (E. Davies, ed.), Spon's Civil Engineering Series, London, 1968.
31. J. E. Baerwald. Traffic Engineering Handbook. Institute of Traffic Engineers, 1965.
32. P. R. Stopher and A. H. Meyburg. Urban Transportation Modelling and Planning. Lexington Books, Lexington, MA, 1975.
33. S. L. Cohen and G. Euler. Signal Cycle Length and Fuel Consumption and Emissions. TRB, Transportation Research Record 667, 1978, pp. 41-48.

Publication of this paper sponsored by Committee on Urban System Operations.