Route Control – Simulation Experiments

M. J. MAHER and R. AKÇELIK

REFERENCE:

NOTE:
This paper is related to the intersection analysis methodology used in the SIDRA INTERSECTION software. Since the publication of this paper, many related aspects of the traffic model have been further developed in later versions of SIDRA INTERSECTION. Though some aspects of this paper may be outdated, this reprint is provided as a record of important aspects of the SIDRA INTERSECTION software, and in order to promote software assessment and further research.
ROUTE CONTROL—SIMULATION EXPERIMENTS

M. J. MAHER
Department of Probability and Statistics, Sheffield University, Sheffield S3 TRH, England

and

R. AKCELIK
Engineering Division, National Capital Development Commission, Canberra, A.C.T., Australia

(Received 2 July 1975; in revised form 30 April 1976)

Abstract—The paper investigates the potential savings in delay likely to result from the use of route control techniques in an urban network of streets where the dominant control mechanism is traffic signal control. The distinction is drawn between user- and system-optimizing solutions to the minimum journey time problem with trip end constraints. Three networks and five origin-destination flow patterns are used in simulation experiments using a stochastic, semi-macroscopic traffic model, a partial loading assignment model and a free-building signal optimization algorithm. The results of these experiments show that substantial savings in average total journey time (between 0 and 25%) could be obtained by the use of route control rather than area traffic control alone.

INTRODUCTION

The review of previous work on route control (Akçelik and Maher, 1976) has helped to define the most relevant context in which to investigate the potential advantages of route control. That context is an urban network of roads in which the dominant control mechanism is traffic signals or, more specifically, an area traffic control system. The aim of route control is, given the network structure and the origin-destination flow matrix, to spread the load throughout the network, so as to optimise some network performance index, such as total network travel time.

Without route control, the signal plan optimisation and assignment process will work together to produce a “user-optimising solution”, which is stable in the sense that no driver can reduce his journey time by changing his route. It should be emphasised that this solution will, in general, be quite different from the “system-optimising solution”, in which the total network travel time is minimised. There are three distinct means of obtaining this desired flow pattern: (i) traffic signal control, (ii) traffic management measures (such as right-turn bans or one-way streets), and (iii) route control devices (such as advisory or mandatory route signs). The third of these is likely to give rise to the most practical difficulties because of the uncertainty of the response to an advisory sign and the enforcement of a mandatory sign. A possible alternative, therefore, to a complete system-optimising solution, is a near-optimal stable distribution using traffic signal control and traffic management measures. As Potts and Oliver (1972) point out, there may be many user-optimising solutions, but there is only one system-optimising solution. It will not, in general, be an easy matter to find the best stable solution, but one obvious approach is to find the overall optimum and search in that neighbourhood for a stable solution. Circumstances and technology dictate whether an overall optimum can be applied in practice, but whether it can or not, the primary objective is to find this system optimising solution. The problem is then: given a network structure and a flow matrix, find the flow pattern and signal plan which will optimise the network performance index.

SIMULATION MODEL

In order to try out the ideas mentioned above and to investigate the route control problem, it was necessary to use a simulation model which was sufficiently sophisticated to demonstrate the type of effects which were thought to be important in route control. The first decision concerned the traffic model: should it be stochastic or deterministic, microscopic or macroscopic, and should it be periodic-scanning or event-scanning? In the light of previous simulation work the model was chosen to be stochastic, semi-macroscopic and event-scanning, using the simulation language SIMON, based on ALGOL. Arrivals at the input points to the network were Poisson, and link travel times were taken from a uniform probability distribution with a standard deviation to mean ratio of 0.289. This was determined by calibration of the model against the results of Ramsey and Hartley (1972) for transformed normal, geometric and rectangular distributions. The mean link travel time was allowed to vary with link flow rate, to simulate the delays due to congestion. The form of this relationship was determined by a car-following model of the type described and analysed by May and Keller (1967). The steady-state speed-flow equation was transformed into a running time-flow equation, and was then approximated by the piecewise linear relationships:

\[
\begin{align*}
\bar{t}/\bar{t}_f = & \begin{cases} \\
1.0 - 0.2y (0 \leq y \leq 0.6) \end{cases} \\
& \begin{cases} 0.15 + 1.6y (0.6 < y \leq 1.0) \\
-48 + 50y (y > 1.0) \end{cases}
\end{align*}
\]

where \(\bar{t}\) is the mean running time, \(\bar{t}_f\) is the free-flow mean running time and \(y\) is the ratio of flow to saturation flow.
The discharge from the stop lines was assumed to be at a constant rate of 800 vph per lane (a constant headway of 2 sec) for non-right-turners. Right-turners were subject to gap acceptance when there was an opposing flow, with a minimum headway of 3 sec, and discharge during the intergreen period was also allowed for. The gap acceptance function was a step function with a critical gap of 5 sec (a trapezoidal function was used at an early stage in the tests, but the difference in the relevant traffic statistics was insignificant).

The signal optimisation programme was developed during this research work and the offset selection procedure was a modification of the Volume Priority method described by Wagner et al. (1971), combined with the tree-building algorithm of Inose et al. (1967). The cycle time and green splits are calculated for each intersection by the method of Webster (1958) and the maximum cycle time found is then the common cycle time for the network. Maximum and minimum values for the cycle time were set at 120 and 40 sec, whilst the minimum green was 10 sec. The whole method has the considerable advantage of speed and simplicity, because (i) the tree-building algorithm solves the problem of closed loops in the network efficiently and (ii) the ideal offset for a platoon moving from link i to link j is taken to be \( t + (g_i - g_j)/2 \), where \( g_j, g_j \) are the green periods upstream and downstream. The testing and validation of this signal optimisation procedure was carried out in a series of experiments to be described later.

The final important component of the simulation model was that of assignment. In order to achieve the accuracy which was felt to be required in the route control experiments, a partial loading technique was used. This was a modification of that described by Steel (1965), and consisted of an incremental, or partial loading process using eight stages. The network was loaded with a fraction of the traffic, the model was run and the link flows noted. These link flows were used to determine the mean link travel times to be used in the next stage of the loading process. After eight stages, all the traffic had been loaded and a final run of the model gave all the necessary traffic statistics. The shortest route algorithm used was based on that of Kirby (1965), which allows for turn penalties and prohibitions and has the "once-through" property.

Further details of the traffic, signal optimisation and assignment models, and the programme organisation are given in Akçelik (1974).

VALIDATION TESTS

The first and most important tests to carry out with a new simulation model are validation tests: tests of the efficiency and accuracy of the model. These will be described only briefly, as they are not the main purpose of this paper, and details are given for those interested, by Akçelik. Three of the four tests were comparisons with previous workers' simulation results.

1. Isolated intersection. The model was run to simulate the delays to a single flow of vehicles arriving at an isolated set of traffic signals. The flow was Poisson and a comparison of the mean delay per vehicle and other statistics with those of Webster (1958) and Blunden (1971) showed that the model gave very close agreement for all flow values. The simulation speed varied from 15,000/1 for a flow of 50 vehicles per hour, to 750/1 for 855 vehicles per hour.

2. A pair of intersections. The tests were performed here in order to compare the resulting delay-offset relationship with those reported by Rumsey and Hartley (1972), using Pacey's transformed normal, Robertson's displaced geometric and Rumsey's rectangular distributions of travel time between the intersections. The agreement was good. Further tests were carried out, for a variety of values of flow, saturation flow and signal timings, and the offset giving minimum delay was found to be \( t + (g_1 - g_2)/2 \) as mentioned previously and the maximum delay was with an offset of \( t + r_s \), where \( r_s \) is the red time downstream. These results are in agreement with Newell's suggestions (1968), Hillier and Rotth's empirical results (1967), and the method used in SIGOP (see Peat et al., 1968). ("Offset" is defined here as the difference in time between the start of the two green periods.) Simulation speeds were in the range of 248/1 to 670/1.

3. Three intersections. Tests were done for a set of three intersections in series and for a single flow passing through them. The results for the delay-offset relationships were compared with those of Watjen (1965). The Watjen model assumed constant journey times and so it is to be expected that there would be some difference due to effect of platoon dispersion. For the purpose of comparison the model was changed to one with constant journey times, and the agreement was then quite close. One result of some interest from these tests was that the delay on the second link depended not only on the offset on that link but also on the offset of the previous link. Simulation speeds were in the range of 325/1 to 355/1 for tests using the rectangular distribution, and from 330/1 to 400/1 for those in which the journey times were constant.

4. A closed network of four intersections. In this test, the network consisted of four intersections, six internal links and six input links. All types of turning movement were involved, with a maximum of 30% left-turning and 28% right-turning proportions on any link. Flow rates were of medium value with degrees of saturation between 0.15 and 0.44. The simulation speed was 691/1, a 3 hr run plus 10 min fill time took 165 sec on the Leeds University 1906A computer, and required 19 K of storage. The differences between expected and measured flow rates were in the range ±3%, and the maximum difference between the expected and measured turning percentages on any link was 2%.

The conclusion was, therefore, that the simulation model behaved efficiently and with sufficient accuracy to be used with confidence in the route control experiments.

ROUTE CONTROL EXPERIMENTS

The aim of these experiments was to arrange the network structure and flow matrix so that there were some flows which had a choice of route. The size of the network and the number of such flows to be kept reasonably small so that it would be possible to trace the causes of whatever effects were noticed, and also to keep
the simulation speed high enough to keep within reasonable bounds the computing time required. The experiments were carried out with steadily increasing complexity of the network, starting from one of the simplest networks which would give some degree of route choice. Throughout these experiments the signal control and route control policies were fixed-time.

The patterns of all the experiments are the same: the stable user-optimising solution is found first of all, by successive applications of the signal optimisation and assignment procedures until equilibrium is reached (this has been described in some detail by Maher and Akçelik (1975)). It is possible for this iterative process either to improve or to worsen the network performance. Secondly, the system-optimising solution is found by an iterative search procedure, using the assignment routine to establish the necessary direction of changes in the link flows in order to obtain an improvement in the network performance, and then computing a new signal plan for the modified flows.

**NETWORK 1**

In this experiment, details of which are shown in Fig. 1 and Table 1, there is a heavy flow of 2000 vph southwards through intersections 1 and 3 and a flow of 600 vph westwards from C to X. The other two flows, q_21 and q_6, have a choice between two routes, 2-1-3 or 2-3.

<table>
<thead>
<tr>
<th>From</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>2000</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>600</td>
</tr>
<tr>
<td>C</td>
<td>600</td>
<td>600</td>
</tr>
</tbody>
</table>

In the user-optimising solution, all of q_21 uses the route 2-1-3 (route 1), but only 70% of q_6 uses that route, leaving 180 vph to travel on the direct route 2-3 (route 2). The network performance, measured by the total network travel time, in vehicle-hours per hr, Z, is 60.8. The system-optimising solution gives a minimum value for Z of 33.7 vehicle-hours per hr by sending all of q_6 by route 2-3 (which could be achieved by the banning of right-turns at intersection 2). Figures 2 and 3 show the results in more detail, giving the relationship between Z, t_1, t_3, and f, where t_1 and t_3 are the average travel times by the alternative routes and f is the proportion of q_6 travelling on route 1. Figure 2 applies when the signal plan used is the one which gives the user optimising solution. It can be seen that the t_1 and t_3 curves intersect at f = 0.7, so that the equal travel time principle of assignment holds. Figure 3 on the other hand, shows the results which arise when the system optimising signal plan is used. The optimum solution Z_max occurs at an end point. The reduction in Z is approximately 12%, and, for the
individual flows, the average journey time for \( q_{av} \) has decreased from 88 to 85 sec, and for \( q_{av} \) from 71 to 60 sec. Although the gains are not uniformly distributed, therefore, no flow suffers an increase in journey time.

A second set of experiments were carried out using network 1, but with a slightly modified demand pattern: a flow of 900 vph arriving at intersection 3 from the west, is assumed to share the same signal phase with flow on link 2–3. The green fraction allocated to the flow on link 1–3 is consequently decreased, increasing the delay on that link whilst providing a larger spare capacity on link 2–3. This is now a two-route, two demand problem unlike the previous single demand problem, because flow \( q_{cv} \) may now be able to use link 2–3 in addition to flow \( q_{av} \). The same procedure as that described previously produced the following solutions:

<table>
<thead>
<tr>
<th>Solution</th>
<th>Flow, vph (av. travel time, sec)</th>
<th>Flow, vph (av. travel time, sec)</th>
<th>Total Network travel time (veh-hrs/hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>User-optimising</td>
<td>300 (99)</td>
<td>300 (99)</td>
<td>72.3 (99)</td>
</tr>
<tr>
<td>System-optimising</td>
<td>0 (79)</td>
<td>600 (83)</td>
<td>57.3 (99)</td>
</tr>
<tr>
<td>Practical</td>
<td>0 (79)</td>
<td>600 (93)</td>
<td>61.6 (99)</td>
</tr>
</tbody>
</table>

The overall minimum \( Z \) is 21% lower than that given by the user-optimising solution. However, to achieve this, it is necessary for 50% of \( q_{cv} \) to travel by each of the two routes. It might be easier to achieve the third solution listed above which allows all of \( q_{cv} \) to travel by route 1, at a cost of an increase of 6% in \( Z \). In both solutions a right-turn ban at intersection 2 must be used, to force all of \( q_{cv} \) to use route 2.

The larger the network, the greater, in general, will be the choice of routes and the greater the scope for route control advantages to show themselves. The next set of experiments use a slightly larger network and it will be seen that the search for the system-optimising solution is a little more complex.

The network used is shown in Fig. 5 with the flow matrix in Table 3.

**NETWORK 2**

This series of experiments used the network shown in Fig. 4 and the demands shown in Table 2.

Flows \( q_{av} \) and \( q_{cv} \) both have a choice between route 1 and route 2 (1–2–3) and route 2 (1–3), but \( q_{av} \) will, in fact, always use route 2. For \( q_{cv} \) the choice rests largely on the right-turning delays at either intersections 1 and 3 or at intersection 2. The free-flow travel times are identical for the two routes.

This example is interesting because it turns out that the user-optimising and system-optimising solutions are identical, both requiring 50% of \( q_{av} \) to use each route and all of \( q_{av} \) to use route 2. There is, therefore, no saving in using route control over signal control. In addition, the network performance curve is almost flat over the range \( 0 < f_{av} < 0.6 \), where \( f_{av} \) is the proportion of \( q_{av} \) using route 1.

**NETWORK 3**

The experiments described so far have used small networks where the number of alternative routes is small.

**Table 3. Matrix of flows in vehicles per hr**

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>From</td>
<td>A 0</td>
<td>B 0</td>
<td>C 600</td>
</tr>
<tr>
<td></td>
<td>D 0</td>
<td>A 0</td>
<td>B 0</td>
</tr>
</tbody>
</table>

**Table 2. Matrix of flows in vehicles per hr**

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>From</td>
<td>C 0</td>
<td>A 500</td>
</tr>
<tr>
<td></td>
<td>B 0</td>
<td>C 0</td>
</tr>
<tr>
<td></td>
<td>D 300</td>
<td>B 0</td>
</tr>
</tbody>
</table>

![Fig. 4. Network 2, showing free flow travel times in seconds and saturation flows in vehicles per hour alongside each link.](image)

![Fig. 5. Network 3, showing free flow travel times in seconds and saturation flows in vehicles per hour alongside each link.](image)
This represents a three-demand, three route problem: demand q_{ux} can use 2-1-4-5, 2-3-4-5 or 2-3-5; q_{yx} has the same choice and q_{xy} can use either 3-4-5 or 3-5. The route via intersection 1 will be called route 1, that via link 3-4 route 2, and that via link 3-5 route 3. It was found from the simulation tests that all of q_{yx} uses route 1 and that all of q_{xy} uses route 2, under all conditions, so that the problem reduces to one with a single demand and three routes. The various solutions and resulting average travel times are given below:

<table>
<thead>
<tr>
<th>Solution</th>
<th>Flow, vph (Av. travel time, secs)</th>
<th>Total network travel time (veh-hrs/hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>User-optimising</td>
<td>420 (113) 150 (112)</td>
<td>91.8</td>
</tr>
<tr>
<td>System-optimising</td>
<td>0 (87) 150 (94)</td>
<td>89.4</td>
</tr>
<tr>
<td>Practical</td>
<td>0 (90) 240 (101)</td>
<td>81.9</td>
</tr>
</tbody>
</table>

The system-optimising solution requires that no right turn at intersection 2 should be allowed for the flow q_{ux} and that the flow should be divided into 150 vph turning right at intersection 3 and 450 vph going straight on. The practical solution above recognises that it may not be possible to achieve this split and so it bans the right turn at node 2 but allows a free choice at node 3. The resulting solution (which uses the system-optimising signal plan) is clearly not as good as the overall optimisation but is better than the user-optimising solution. Potential savings in total network travel time are 12% for the overall system-optimum, and the loss due to a change from this to the practical solution is only 25%

Another set of experiments were performed, in which conditions on the arterial road 1-4-5 were worsened by the introduction of two eastbound flows, each of 900 vph, at intersections 4 and 5. Therefore the green times for links 3-4 and 4-5 will be smaller and delays increased, whilst for links 3-4 and 3-5 there will be increased spare capacity and delays will be reduced. As was expected, this change meant that the potential route control savings were increased, as may be seen from the solutions below:

<table>
<thead>
<tr>
<th>Solution</th>
<th>Flow, vph (av. travel time, secs)</th>
<th>Total Network travel time (veh-hrs/hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>User-optimising</td>
<td>270 (129) 150 (129) 180 (129)</td>
<td>110.3</td>
</tr>
<tr>
<td>System-optimising</td>
<td>0 (94) 0 (77) 600 (100)</td>
<td>83.1</td>
</tr>
</tbody>
</table>

The potential savings are approximately 25%, and it should be noted that, in moving from the user-optimising to the system-optimising solution, the average journey times of diverted traffic are decreased (from 129 to 100 sec, and from 94 to 89 sec), even though any single vehicle from q_{ux} would be quicker using either route 1 or route 2 and any single vehicle from q_{yx} would be better using route 2. In this sense, the solution is not stable but requires some route control measures (i.e. control over the route selection decisions of drivers) to preserve the optimality of the solution.

To conclude the description of the simulation experiments, there is one general result which should be mentioned, and that is that the variation in Z, the total network travel time, is almost entirely explained by the variation in the total delays at intersections. This means that the variation in link delays, due to congestion, may safely be ignored in calculating the minimum value of Z. This suggests very strongly that the model could be simplified by omitting the mean travel time—flow relationships without any real loss of accuracy or realism.

Clearly, an even more important simplification of the model would be to use a macroscopic deterministic traffic model (of the TRANSYT type, for instance) instead of the stochastic model used here. Some recent trials have indicated very close agreement between the results from the two types of models and if these are substantiated by further tests, this would obviously give a way of speeding up the whole simulation process and enable larger networks to be studied. The speed of the simulation using network 1 was between 37/1 and 45/1, for network 2 between 70/1 and 76/1, and for network 3 between 18/1 and 23/1.

Another result should be mentioned regarding the potential savings from route control in practice. The savings given above are for a change from the user-optimising solution to the system-optimising one. This is based on the assumption that the user-optimising flow pattern represents the situation found in practice in the absence of route control. However, real-life flow patterns may be different from the user-optimising pattern because of some inaccuracies and irrationalities in drivers' route selection decisions and as a result of non-Wardrop phenomena as discussed by Simoes Pereira (1968). One possibility which would result in larger savings due to a system-optimising route control policy is that the major routes will attract more traffic than in the user-optimising pattern. This may happen because the major routes provide smoother journeys and smaller average delays on individual links, and for similar reasons.

To investigate this possibility for each network described above, a flow pattern in which the travel time on the major route is longer than those on the alternative routes by an amount in the range 10-20% was considered.
This was found as a possible real-life pattern under the user-optimising signal plan. Comparing total network travel time for this pattern with that for the system-optimising solution for each network, it was found that the maximum savings were 19.4 veh-hrs/hr (25%) in network 1, 1.2 veh-hrs/hr (4%) in network 2, and 39.2 veh-hrs/hr (39%) in network 3. Larger savings in average route travel times for both diverted and undiverted vehicles were found in this case. The maximum possible saving in total network delay was 38.7 veh-hrs/hr (50%) found in the second experiment using network 3.

CONCLUSIONS

The purpose of the simulation experiments described in this paper has been to investigate the potential savings to be gained by the use of route control combined with signal control, over signal control alone. The system-optimising solution gives a lower bound on the value of the total network travel time, obtained by treating the flow pattern as well as the signal plan as a control variable. Whether such a solution can be implemented depends on many factors, including the nature of the solution, the technology available and the behaviour of drivers when confronted by advisory or mandatory route signs. The user-optimising solutions calculated here have been those which result from a continuous process of signal plan optimisation and the consequent reassignment of traffic in the network.

The set of experiments, using three different networks, have given rise to various types of results. In the test on network 2, for instance, the user- and system-optimising solutions were found to be identical. Some solutions required the whole of a flow to be diverted on to some new route, whilst other solutions required only part of the flow to be so diverted. The concept of a "stable" solution was introduced, in which no driver could reduce his average journey time by switching to another route. The user-optimising solution is one such stable solution, but it is possible to obtain others, better than that, by the use of a traffic signal plan and traffic management measures. The system-optimising solution would be the most appropriate point to start from in a search for the best stable solution.

It has been found that, generally, in moving from the user-optimising to the system-optimising solution, the average journey times of diverted traffic have been decreased as well as those of undiverted traffic. This clearly is a convenient and advantageous result which would make the success of any such scheme much more likely. The overall saving in total network travel time has been in the range 0-24%, the exact percentage being dependent on the degree of congestion on the major routes and the spare capacity on relatively underused routes.

It has been suggested by Gazis (1971) and Brand (1972) that area traffic control has progressed almost as far as it can, and that the introduction of more sophisticated, dynamic control policies will give only marginal improvements. If this is the case, then route control is the one type of control remaining which can give substantial benefits. The size of these benefits, even in the small networks studied here, has been sufficiently large to demonstrate the need for larger scale investigations. Because of the nature and size of the search required to find the optimum flow pattern, it would be desirable to use a simpler deterministic, macroscopic traffic model for such an investigation using larger networks. Further tests are needed to verify that such a traffic model, combined with an assignment model and a signal plan optimisation model, will fully demonstrate the fundamental route control effects found in the research reported here. It should also be noted that, although the simulation tests reported here were performed on relatively small networks, they may be considered as critical areas of larger networks. This corresponds to the limited use of route control in large networks. The full use of route control in large networks should produce greater savings in delays.

The description in this paper has been traffic oriented, but the authors feel that the concept of route control bridges the gap between what have traditionally been traffic engineering problems (area traffic control schemes, traffic management measures) and transport planning problems (trip distribution, traffic assignment). Apart from having the obvious and often desirable effect of minimising the total network travel time, route control policies would also produce savings in energy consumption, use the existing infrastructure to better effect and delay the expenditure of additional capital on increasing the capacity in the network. The interrelationships between the signal plan and the network flow pattern could also be used to advantage in environmental and traffic restraint schemes.

Acknowledgements—The work described in this paper was carried out whilst M. J. Maher was a member of staff and R. Akçelik a research student in the Institute for Transport Studies and Department of Civil Engineering, Leeds University. The authors wish to thank the University for the extensive use of the computing facilities, and U.N.E.S.C.O. and Kandemir Technical University, Trabzon, Turkey for their financial support of R. Akçelik.

REFERENCES


Brand D. (1972) Urban traffic control: how far can we take it? Traffic Engng 42(11), 50-55.


