Intelligent Control of Urban Road Networks: Algorithms, Systems and Communications

Mike Smith

Department of Mathematics, University of York,
Heslington, York Y010 5DD, United Kingdom
mjs7@york.ac.uk

Abstract. This paper considers control in road networks. Using a simple example based on the well-known Braess network [1] the paper shows that reducing delay for traffic, assuming that the traffic distribution is fixed, may increase delay when travellers change their travel choices in light of changes in control settings and hence delays. It is shown that a similar effect occurs within signal controlled networks. In this case the effect appears at first sight to be much stronger: the overall capacity of a network may be substantially reduced by utilising standard responsive signal control algorithms. In seeking to reduce delays for existing flows, these policies do not allow properly for consequent routeing changes by travellers. Control methods for signal-controlled networks that do take proper account of the reactions of users are suggested; these require further research, development, and careful real-life trials.

Keywords: Intelligent Network Control, Urban Traffic Control, Capacity-maximising Control, Stability of Transport Networks, Complexity.

1 Introduction

1.1 Urban Traffic Control, Healthcare, Computing and Further Research

The problem of controlling or managing a system with many reactive "users" has been considered within studies of urban traffic control for many years; various difficulties have arisen and various solutions have been proposed. Similar difficulties arise in other fields; for example healthcare and computing both involve networks of facilities with capacity and budget constraints and many "users" who make their own decisions.

In this paper we consider only the traffic control case hoping that this case may yield insights in these other areas. In any case the author knows only about this case!

This paper considers several simple examples all designed to show that feedback loops can have a significant effect on outcomes of utilizing apparently sensible control policies or algorithms. Some examples show that other novel policies may perform much better than standard policies; because they deal with feedback better.

Novel control policies discussed here merit further research within the transport field. These policies may also be useful when healthcare and computer networks are to be managed and controlled; and so research on transferring these approaches to

R. Mehmood et al. (Eds.): EuropeComm 2009, LNICST 16, pp. 116–127, 2009.

both those fields is needed. Finally, while we have considered stability briefly in this paper, speed of reaction has not been considered in any detail. Yet speeds are now increasing rapidly in real life in many vital systems. Stability when communications and reactions are very fast merits much more attention in many fields including transport networks.

1.2 Urban Traffic Control and Traffic Assignment and the Digital Economy

Given an urban network with a given set of signal controls, a control or management change designed to benefit network users may actually make network users worse off; due to "knock on" effects which have not been properly allowed for. These knock-on effects arise as a result of the reactions of network users who are free to change their route and mode; the flow pattern can and does change in reaction to control changes. Beckmann once said: "travellers will be playing their own games while the network manager is playing his".

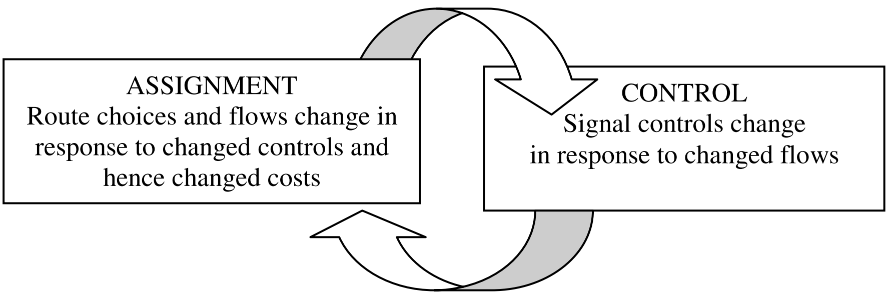

Fig. 1. The assignment–control interaction loop above is traversed indefinitely; the loop represents the interaction between a control system (where signal controls change in response to changed flows on the network according to some policy or algorithm) and an assignment system (where route-flows change toward cheaper routes as drivers respond to changed control settings and hence changed costs)

If in the loop in figure 1 traffic signal control settings are changed to achieve some aim (such as reducing the average journey time of travellers on the assumption that flows are fixed) then the following (or knock-on) change in the routes chosen by drivers (moving to cheaper routes say) may cause delays to increase instead.

Allsop in [2] first stated that route choices need to be taken into account when optimising signals and Dickson [3] gave an example to show that continually optimising signals assuming flows are fixed does not achieve the optimum result when the flows are, as is usually the case, variable.

There is as yet no reliable automatic procedure for finding (for example) a change in the control parameters of an urban traffic control system which is certain to benefit the network as a whole when the responses of divers are taken into account. This problem has been much considered; see, for example, [4, 5, 6]. Much more mathematical / computing research is needed here before real effects are felt on-street.

The digital economy is moving this problem (in both the economic sphere and the transport sphere) into a whole new era of very fast communication. *The difficulty of taking proper account of knock-on effects is multiplied when this has to be done fast.*

The speed of onset of the credit crunch is a symptom of this. Movements of international capital, reacting perhaps to excessive lending in the USA and elsewhere, have left governments to "catch up", using laborious regulatory tools, aiming to control rapidly reactive international knock-on effects. The computing power available in the digital age has improved the tools available to managers, but it has also changed forever the "game" which managers are seeking to manage or control; with uncertain consequences.

2 Considering Braess's Network

Assume first that the dotted link is not present. A total flow of 6 travels from the origin to the destination on the network of four bold-arrow links. The two route costs are equal and so no element of flow has any incentive to change route if

flow on route 1 = flow on route 2 = 3.

The average travel cost at this equilibrium is

cost of route 1 = cost of route 2 = (10.3) + (3 + 50) = 83 units.

Now assume that the dotted link has been added. If k = 10 (as Braess originally specified) all three routes have the same cost if

flow on route 1 = flow on route 2 = flow on route 3 = 2.

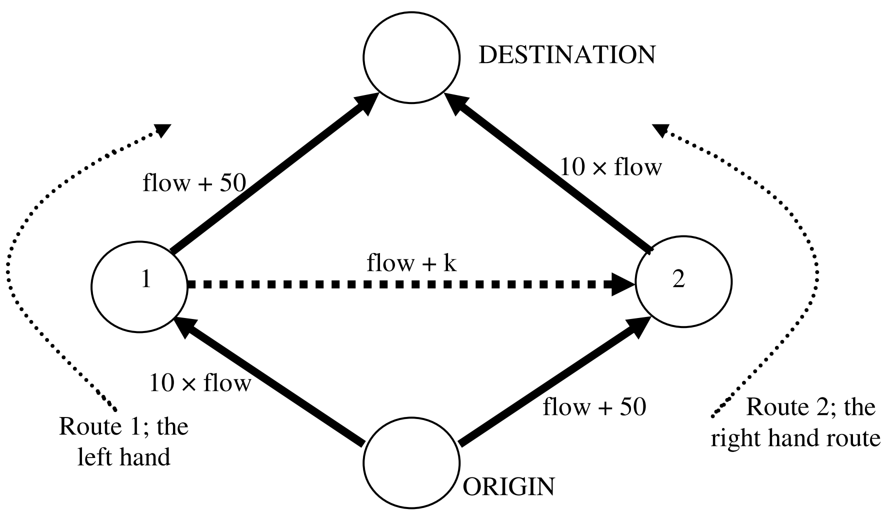

Fig. 2. The network first considered by Braess [1] to illustrate the counterintuitive effects which may arise when a network link is added to reduce congestion. Costs of traversing the links are shown: 10 × flow means that the link cost equals 10 times the flow on that link, and so on. Braess considered the effect of adding the dotted link to the basic network of four links indicated by solid arrows. See [7] for an English translation of Braess's 1968 paper.

The average travel cost at this equilibrium distribution (which we write as (2, 2, 2)) is now 10.(2+2) + 2 + 50 = 92 > 83; this is the travel cost of each route when the traffic distribution is (2, 2, 2).

In the augmented network, with the added dotted link and where k = 10, (3, 3, 0) is the optimal distribution of the total flow of 6; this distribution minimizes average or total travel cost. So the minimum average cost for this new augmented network is 83.

For the augmented network it is natural to consider the ratio of the equilibrium cost and the optimal cost, and Roughgarden [8] has called this ratio the price of anarchy. The price of anarchy for the augmented version of the Braess network is thus 92/83. Roughgarden [8] shows that if all cost functions have the form (a × flow) + b then (total cost at equilibrium flow pattern)/(total cost at optimum flow pattern) ≤ 4/3.

Valiant and Roughgarden [9] show that for large random networks (with link cost functions of the form (a × flow) + b) the ratio above approaches 4/3 with probability 1 as the number of links in the network tends to infinity; provided network loads are suitably (adversarially) chosen.

2.1 Controlling Braess's Augmented Network by Varying k in Figure 2

Consider again the augmented version of the Braess network. Consider letting k = 23 (instead of k = 10, the value chosen by Braess). It is easy to see that the flow distribution (3, 3, 0) now becomes an equilibrium with the costs of all three routes being the optimal 83. Removal of the dotted road is unnecessary; just increase its cost.

The equilibrium network performance changes as k varies in figure 2: k = 10 gives the Braess network with equilibrium (2, 2, 2) while k = 23 ensures that the dotted link is sufficiently unattractive to be unused at equilibrium and the equilibrium then becomes (3, 3, 0); as if the dotted link had indeed been removed. Let k be a "control"; if 0 ≤ k ≤ 23 then (2 + (k-10)/13, 2 + (k-10)/13, 2 – 2(k-10)/13) is the unique equilibrium flow pattern: all route costs equal 92 – 9(k-10)/13, as shown in figure 3.

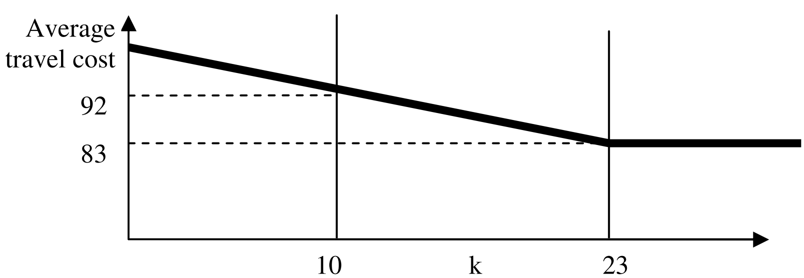

Fig. 3. The performance of the augmented Braess network versus k for non-negative k. Average travel cost at equilibrium is a decreasing function of k for 0 ≤ k ≤ 23.

3 Network Control Using Signals

In figure 4, two routes join a single Origin – Destination pair. There are two stages at the signal; the first gives green (or right of way) to route 1 and the second gives green (or right of way) to route 2. We envisage that this network is traversed day after day and that some drivers change route to a cheaper route if a cheaper route is available.

We show that changing the control policy from a standard policy to the P_0 policy (introduced in [10, 11, 12]) doubles the capacity of this network.

So suppose first that the drivers on the network in figure 4 have adjusted their routes over time so that both routes currently have equal cost; then, since route 1 is less costly than route 2 (ignoring the delays), the delay at the signal on route 1 must exceed the delay at the signal on route 2. Suppose also that flow is currently greater on route 1.

As both the flow and delay at the signal are greater on route 1 than on route 2, to reduce current delay some green-time must be swapped from the upper route to the lower route. Suppose some green-time is swapped.

This green-time swap will increase delay at the signal on route 2 and reduce delay at the signal on route 1. There is now an incentive for drivers to swap from route 2 to route 1. Suppose that some small proportion of drivers on route 2 do swap to route 1.

We are now in a similar position to that in italics two paragraphs up; and so the same argument works again (and again and again . . . .). Eventually all traffic and all available green time will become allocated to route 1. Since route 2 is twice as wide as route 1 this halves the maximum possible capacity of the network. [Differential equations may be used to make these swapping rules precise. The above argument is largely independent of swap rates and essentially depends only on swap rates being positive.]

The problem of course is similar to that in the Braess network: each green time swap reduces total delay at the signal and reduces total travel costs felt *if the flows remain fixed*. But as flows respond in the natural way to changed delays (by swapping to cheaper routes as above) then total travel costs and delays at the signal in fact increase; just as costs do in the Braess example when k is reduced. (See figure 3.)

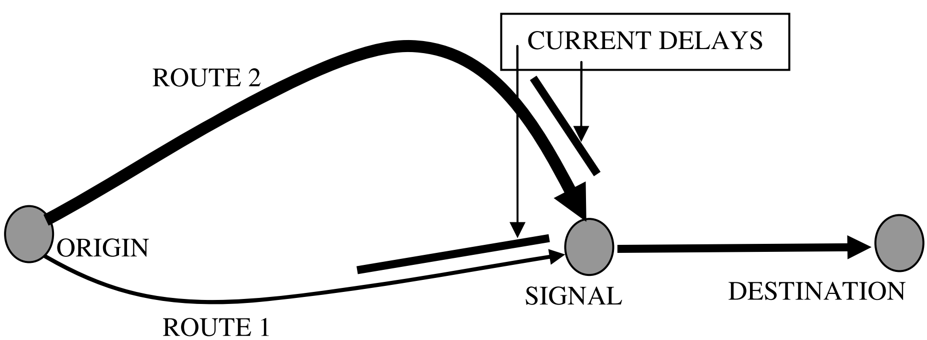

Fig. 4. A simple signal-controlled network. Route 2 is longer and (ignoring the delays at the signal) more costly than route 1. Route 2 is twice as wide as route 1 and the saturation flow at the signal for route 2 is twice the saturation flow at the signal for route 1. The signal responds to the local delays at the signalized junction. The length of each bar represents the delay felt by vehicles at the signal; as shown here these delays are currently larger on the lower route, route 1.

3.1 Designing Control Changes and the Stability of Control Systems

In designing network control changes we need to try to take account of the likely knock-on effects of our actions on future traveller choices.

This is difficult; especially if fast decisions need to be taken as in the traffic control case. Thus it is natural to look for a simple practical solution to this difficulty when control decisions are taken.

One clue is in the last simple example: here not only do delays rise as the swapping processes occur but the overall travel capacity of the network declines as the swapping processes occur. Perhaps if we aim not to minimize delay but to maximize capacity the anticipation problem will become more practically soluble and maybe even soluble *fast*. Maybe even fast enough for application to the design of real time control systems. Furthermore it may be that maximizing capacity will be quite good at reducing delays even if not minimizing delays; we are led to the P_0 control policy.

3.2 A Capacity Maximizing Control Policy or Algorithm: P_0

In this section we specify a signal control policy called P_0. If in the network shown in figure 4 the origin destination flow slowly increases from zero then at equilibrium this policy compares with standard delay minimisation as shown in figure 5 below.

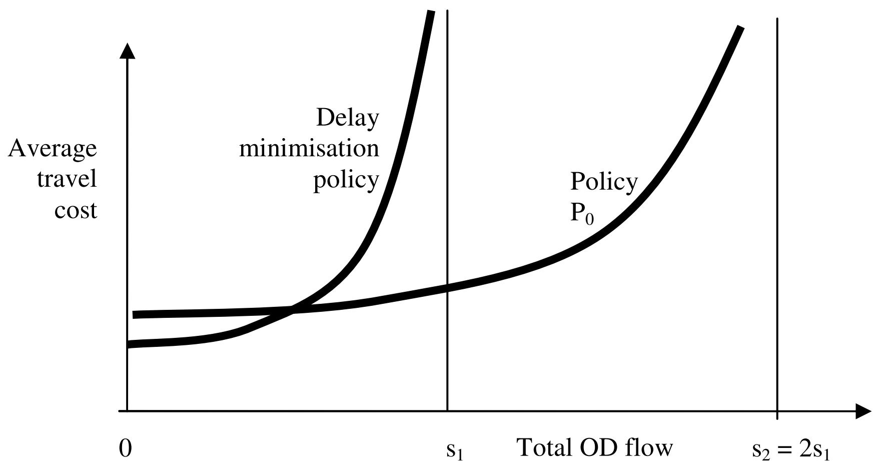

Fig. 5. Equilibrium performances of responsive delay-minimisation and P_0 policies as the total flow from the origin to the destination in figure 4 increases from zero. The capacity of the network is doubled by switching from delay-minimisation to P_0.

To specify the P_0 policy we need a little notation. Suppose that, in figure 4,

s_1 is the saturation flow of route 1 at the junction;
s_2 is the saturation flow of route 2 at the junction;
d_1 is the delay felt at the junction by each vehicle traversing route 1; and
d_2 is the delay felt at the junction by each vehicle traversing route 2.

The delays felt at this signal-controlled junction depend on both flows and green-times and this dependence is here omitted for clarity. The P_0 control policy for a junction like that in figure 4, with just two approaches, is as follows: for given flows and delays at the junction, choose green-times so that $s_1d_1 = s_2d_2$.

In figure 4 we suppose that $s_2 = 2s_1$; the P_0 policy $s_1d_1 = s_2d_2$ thus here ensures that $d_1 = 2d_2$.

We may now justify the P_0 graph in figure 5. Suppose that demand slowly increases from 0. Then initially all traffic will choose route 1 as this is shorter than route 2 and the junction delays are small because flows are small.

Eventually however as the total origin to destination flow increases junction delays will also increase; and then the difference $d_1 - d_2 = d_2 = ½d_1$ also increases. Eventually $d_1 - d_2$ (= d_2) will exceed Δ, the difference between the free-running costs of the two routes. When this happens the travel cost via route 1 will exceed the travel cost via route 2 and travellers will gradually switch to route 2 as the total flow increases. When delays become very high $d_1 - d_2 = d_2$ will also become very high and all travellers will be using the wider route 2, maximising the capacity of the network.

4 More Complicated Systems and Complex Systems

A more complicated interacting system arises if demand is represented, as in figure 6 above. It is then also natural to add more elements into these three boxes: for example

(1) public transport flows may be added into the assignment box to take account of travellers' choice of mode as well as route;

(2) prices and subsidies may be added into the control box to take account of changes in these "controls"; and

(3) bus routes and frequencies may be added into the demand box to take account of changes in these.

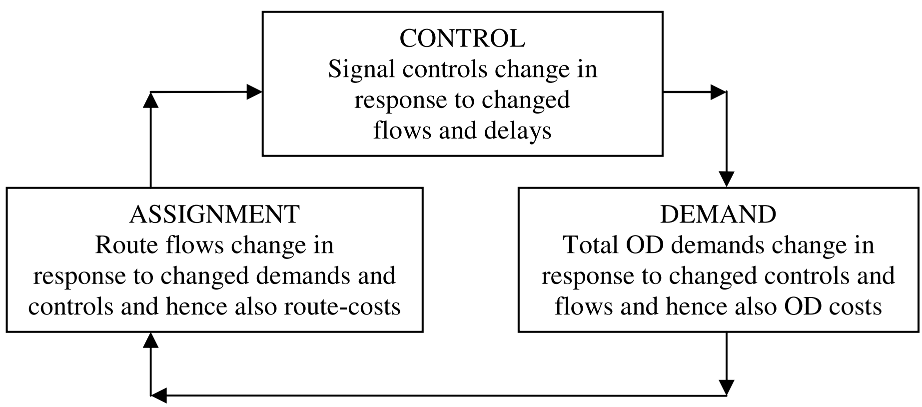

Fig. 6. Assignment–demand–control interaction: figure 1 has been expanded to allow for demand changes. This may represent a real system or a model of a real system.

In both model and real-life dynamical systems the set of equilibria is very important and it is clear that systems like that in figure 6 can get very complicated and then it may be hard to determine the set of equilibria.

However complexity is not just a question of size; as we shall see below.

4.1 A Small Complex System and a Pitchfork

In the signal controlled network shown in figure 7 below, both routes have the same undelayed travel time and the same saturation flow of s vehicles per minute at the signal. We here assume that the signal responds to traffic flows by equalizing saturation ratios.

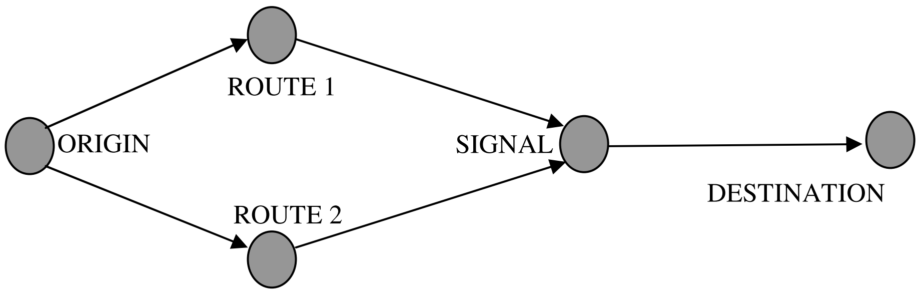

Fig. 7. The signal is adjusted to equalise the degrees of saturation at the signal for any flows along the two routes. The network is symmetrical.

Let T be the total flow from the Origin to the Destination via the two routes and let $0 < T < s$. Let X_1 be the proportion of the Origin-Destination flow which travels along route 1 and X_2 be the proportion travelling along route 2; so that $X_1 + X_2 = 1$. The flow on routes 1 and 2 will then be TX_1 and TX_2. We will consider T fixed but also we will consider different values of T (satisfying $0 < T < s$). Let G_1 be the green time proportion awarded to route 1 and G_2 be the green time proportion awarded to route 2.

Since the signal equalises the two saturation ratios at the signal, X_1, X_2, G_1, and G_2 must satisfy $X_1T/sG_1 = X_2T/sG_2$. Since G_1 and G_2 are green time proportions and so add to 1 it now follows that G_1, G_1 must be given (for any T) by:

$$G_1 = X_1 \text{ and } G_2 = X_2.$$

We further assume that the delay costs d_1 and d_2 felt at the signal by vehicles traversing routes 1 and 2 are determined by putting:

$$d_1 = BTX_1/[sG_1(sG_1 - TX_1)] \text{ and } d_2 = BTX_2/[sG_2(sG_2 - TX_2)];$$

and that the travel costs C_1 and C_2 along routes 1 and 2 are then given as follows:

$$C_1 = ATX_1 + d_1 \text{ and } C_2 = ATX_2 + d_2.$$

A and B are constants; delay cost here is identical to the second term of Webster's delay formula [13] if B is chosen to be 9/20; in this case d_1 and d_2 will be the estimated delays in minutes per vehicle.

Since G = X it follows immediately that

$$C_1 - C_2 = AT(X_1 - X_2) + [BT/s(s-T)][1/X_1 - 1/X_2].$$

To find equilibria; where route costs are equal and so no traveller has any incentive to change route; we need to solve the equation $C_1 - C_2 = 0$. For any T, $X_1 = X_2 = ½$ yields one solution and so is an equilibrium: are there others?

Multiply the equation $C_1 - C_2 = 0$ through by X_1X_2 and consider instead:

$$(C_1 - C_2)X_1X_2 = AT(X_1 - X_2)(X_1X_2) - [BT/s(s-T)][X_1 - X_2] = 0.$$

Now either $X_1 - X_2 = 0$ or (dividing by $X_1 - X_2$ and letting $X_2 = 1 - X_1$):

$$X_1^2 - X_1 + B/[As(s-T)] = 0.$$

Hence

$$X_1 = ½ + [¼ - B/As(s-T)]^{1/2} \text{ or } X_1 = ½ - [¼ - B/As(s-T)]^{1/2}.$$

There are two additional *real* roots of the quadratic equation (and so of $C_1 - C_2 = 0$) and so two additional equilibria if and only if $¼ - B/As(s-T) > 0$ or $T < s - 4B/As$.

The complete set of equilibrium (TX_1, TX_2) as T varies from 0 to s is shown by bold lines in figure 8. [Each "axis" of the figure also consists of equilibria, and so these are bold too.]

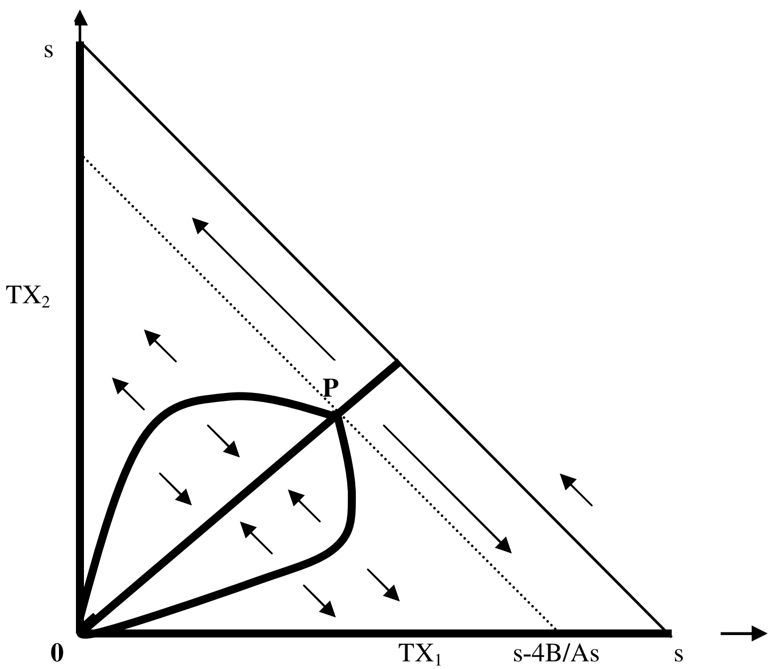

Fig. 8. Bold lines indicate the set of route-flows TX = (TX_1, TX_2) which are equilibria consistent with the equisaturation policy for some T. Five branches of the equilibrium set, three of them comprising stable equilibria, converge on the origin as T becomes small. Arrows indicate the natural directions of motion of non-equilibria (for various values of T) as drivers and green-time respond to delays and saturation ratios. If X_1 is plotted against T the shape shown here becomes a pitchfork.

Let us suppose that T slowly increases from zero. If the flow pattern initially follows the central (stable) equilibrium set leading away from the origin then it will naturally carry on doing that until T = s – 4B/As, when the bifurcation point P is reached. As T increases beyond this point the central equilibria become unstable and the flow pattern will then naturally follow one of the long arrows and continually move toward one of the two axes.

Thus the behaviour of the system changes sharply near P. This is in part due to the change in the topological structure of the equilibrium set at this point. Consider a small circle around P: for T < s – 4B/As the equilibrium set has three points within the circle and for T ≥ s – 4B/As the equilibrium set has just one point within the circle. Since the topological structure of the equilibrium set changes at P this point is called a *bifurcation.*

4.2 Real Life Control Systems and the Origins of Traffic Assignment

An early signal setting method for an isolated traffic signal (essentially the equisaturation method) was specified and justified by Webster [13]. TRANSYT (Robertson [14]) is a method of designing fixed time signal timings in a network.

Many cities use the traffic-responsive SCOOT control system; see Hunt [15].

Wardrop [16] introduced the equilibrium condition discussed in this paper. This is: at equilibrium, no driver has a less costly alternative route.

5 Conclusion

This paper considers control in road networks. Using a simple example the paper has shown that reducing delay on certain network links may increase delay when travellers react by changing their travel choices; and increasing delay on certain links, encouraging a more efficient equilibrium to arise, may reduce delay when travellers change their travel choices. This example is based on the Braess network [1].

Valiant and Roughgarden [9] and others suggest that Braess effects may happen in many networks; specific examples are given by Cohen and Kelly [17] and in [10].

It is shown in this paper that the Braess effect can occur in a powerful form within signal controlled networks. The overall capacity of a small network is substantially reduced by utilizing standard signal control algorithms, especially if these are used responsively. The paper has also shown that, on the other hand, the P_0 policy maximizes the capacity of this small network; Smith [12] shows that, under natural conditions, this capacity-maximising property holds in a general signal controlled network.

Stability is a vital element of any system and has been considered only very briefly here. Some general stability results have been given in Smith and van Vuren [18]. Of the six policies or algorithms considered there, the most clearly stable (allowing for the possibility of routeing changes) are all similar to P_0.

5.1 Further Work on Capacity Maximizing Policies and Bilevel Programming

These results suggest that further study of capacity-maximising control policies, embracing (1) mathematical study, (2) computer tests on larger scale model networks, (3) dynamical considerations and (4) real-life tests may all be worthwhile.

Transfer of the capacity-maximizing ideas to other fields, including perhaps healthcare networks and computer systems and the internet, also merits attention.

Finally there is a great need to extend the work described in [4, 5, 6] on bilevel optimization; and then to construct really effective computer programs optimising general networks subject to choices by stakeholders. This is certainly a grand challenge.

Acknowledgment. The author is a Researcher/Co-investigator on the FREEFLOW project. FREEFLOW is aimed at designing and implementing systems which transform road transport data into intelligence; to help travellers, operators and network managers make better (and more timely) decisions. FREEFLOW is funded by the Technology Strategy Board, the Department for Transport, the EPSRC and the partners. These partners are: Transport for London, City of York Council, Kent County Council, ACIS, Kizoom, Mindsheet, QinetiQ, Trakm8, Imperial College London, Loughborough University and the University of York. Ian Routledge is a subconsultant to the University of York and the Technical Director of FREEFLOW is Andy Graham (White Willow Consulting).

References

1. Braess, D.: Uber ein Paradoxon aus der Verkehrsplanung. Unternemensforschung 12, 258–268 (1968); English translation in [7] below
2. Allsop, R.E.: Some possibilities for using Traffic Control to Influence Trip Distribution and Route Choice. In: Proceedings of the Sixth International Symposium on Transportation and Traffic Theory, pp. 345–374. Elsevier, New York (1974)
3. Dickson, T.J.: A note on traffic assignment and signal timings in a signal-controlled road network. Transportation Research B, 267–271 (1981)
4. Smith, M.J.: Bilevel optimisation of prices and signals in Transportation Models. In: Lawphongpanich, S., Hearn, D.W., Smith, M.J. (eds.) Mathematical and Computational Models for Congestion Charging, pp. 159–199 (2006)
5. Luo, Z.Q., Pang, J.S., Ralph, D.: Mathematical programs with equilibrium constraints. Cambridge University Press, Cambridge (1996)
6. Fletcher, R., Leyffer, S.: Nonlinear programming without a penalty function. University of Dundee Numerical Analysis report NA 171 (2000)
7. Braess, D., Nagurney, A., Wakolbinger, T.: On a Paradox of Traffic Planning. Transportation Science 39(4), 446–450 (2005)
8. Roughgarden, T.: Selfish routing and the price of anarchy. MIT Press, Cambridge (2005)
9. Valiant, G., Roughgarden, T.: Braess's Paradox in Large Random Graphs. In: Proceedings of the 7th ACM Conference on Electronic Commerce (2000)
10. Smith, M.J.: A local traffic control policy which automatically maximises the overall travel capacity of an urban road network. In: Proceedings of the International Conference on Urban Traffic Control Systems, Berkeley, California (August 1979); Traffic Engineering and Control 21, 298–302 (1980)
11. Smith, M.J.: Traffic control and route choice; a simple example. Transportation Research 13B, 289–294 (1979)
12. Smith, M.J.: The existence, uniqueness and stability of traffic equilibria. Transportation Research 13B, 295–304 (1979)

13. Webster, F.V.: Traffic signal settings, Department of Transport, Road Research Technical Paper No. 39, HMSO, London (1958)
14. Robertson, D.I.: TRANSYT: a traffic network study tool. RRL Lab. report LR253, Road Research Laboratory, Crowthorne, UK (1969)
15. Hunt, P.B., Robertson, D.I., Bretherton, R.D., Winton, R.I.: SCOOT - A Traffic Responsive Method of Coordinating Signals. Transport and Road Research Laboratory Report 1014, Transport and Road Research Laboratory, Crowthorne, UK (1981)
16. Wardrop, J.G.: Some theoretical aspects of road traffic research. In: Proceedings of the Institute of Civil Engineers, Part II, vol. 1, pp. 325–378 (1952)
17. Cohen, J.E., Kelly, F.P.: A paradox of congestion in a queueing network. J. Appl. Prob. 27, 730–734 (1990)
18. Smith, M.J., van Vuren, T.: Traffic equilibrium with responsive traffic control. Transportation Science 27, 118–132 (1993)