Upstream-gating merge-control which maximises urban road network capacity and reduces bus delays

Emeritus Professor Michael J Smith [a], Dr. Richard Mounce [b]

[a] Department of Mathematics, University of York, UK
[b]RBM Traffic Solutions Ltd, 95 Princess Drive, York, YO26 5SX, UK.

Email-addresses: mike.smith@york.ac.uk (M. J. Smith), richard@rbmtrafficsolutions.com (R. Mounce)

ABSTRACT

This paper considers responsive signal control and suggests a new approach to the design of responsive signal control policies. The paper suggests that responsive control should have a basic aim of *maximising network capacity* (assuming that route choices are variable); instead of aiming to minimise delays. The paper then illustrates several capacity-maximising control policies using two simple networks, and briefly considers their effects on buses.

With the first network the paper states a policy which is not capacity-maximising on this network and gives rise to ever-increasing queues. On the other hand, the paper defines control policy P_0 and shows that this policy is capacity-maximising for this network.

The second network is a simple network with a merge and a downstream bottleneck, which corresponds to a very congested part of the City of York road network - Gillygate. The paper focusses on "upstream-gating" control strategies which hold traffic back at traffic signals just ahead of the merge to prevent the formation of a queue at the downstream bottleneck; and also considers different ways of further, additionally, controlling the two inflows to the merge. It is shown that adding upstream merge-control which follows P_0 maximises the network capacity, notwithstanding the upstream gating. This is supported by

(i) a dynamic analysis allowing for the dynamic growth of queues in the second network and
(ii) real-life results of upstream-gating applied to Gillygate in York (UK) which provided motivation for this paper; these results show the potential value of upstream gating in real life.

Based on the results here and other results in other papers this paper suggests that upstream-gating merge-control which maximises network capacity should be considered as a new basic responsive signal control strategy.

Keywords: Upstream gating; merge control; capacity-maximising signal control.

1. Introduction

1.1. The impacts of motor traffic in towns

The distribution of motor vehicle traffic flows and queues both have very important impacts on the effective use of urban road-space. For example:

(i) active travel is much more likely to thrive where motor vehicle flows and queues are small,
(ii) public transport runs more efficiently where motor vehicle flows and queues are small,
(iii) enjoying the view of inspiring architecture is often much enhanced by reductions in the visual intrusion of motor vehicles, and
(iv) the use of streets for play and other social interactions becomes possible when traffic flows and queues are small or very small.

Traffic signals may be regarded as taps controlling both flows and queues. It follows that the interaction between traffic signal timings and motor traffic flows and queues on a road network is key to many important aspects of urban traffic. For example, if specific limits for *motor traffic flows* and the *number of motor vehicles* on each street are agreed then signal timings may be used to help achieve these *flow limits* and *vehicle number limits*; as well as maximising network capacity.

1.2. The interaction between traffic signal control and flows and queues of motor traffic

Allsop (1974) suggested that traffic signal-settings should take account of drivers' route choices; so as to beneficially affect the distribution of motor vehicle traffic flows and queues on an urban road network. This paper is concerned with responsive or adaptive signal controls where the green-times are responsive to traffic flows and queues; the paper addresses the question:

how should green-times respond to flows and queues?

The paper makes several suggestions in response to this question, taking careful account of drivers' route choices.

Motor vehicle traffic congestion, and resulting queues on road networks, represent a high cost for mobility by car, bus, bicycle and walking, including wasting fuel, causing stress for all road users and generating emissions that impact health and climate change.

This paper focusses on developing new basic responsive traffic signal control strategies which may be a much better basis for meeting both old and new objectives than existing responsive traffic signal control strategies.

Now is an appropriate time to develop new more helpful signal control strategies; because

(i) it is increasingly clear that current signal control strategies do not meet current needs,
(ii) current needs are changing, and
(iii) it is now becoming possible to utilise vehicle travel times and delays to adjust signal controls.

1.3. Measures of performance and route choice assumptions

The classic initial or base performance measures adopted to guide the design and implementation of fixed time and responsive traffic control systems have been:

travel times or delays experienced by vehicles and *number of stops experienced by vehicles.*

In particular, TRANSYT (Robertson, 1969) and SCOOT (Robertson and Bretherton, 1991) were initially designed to

minimise [(total delay) + w×(total number of stops)]
on the assumption that (1)
route choices are fixed.

There have since been many additions and amendments to SCOOT (see for example Bretherton et al (2002)); to take better account of many aspects of real-life traffic. These include: bus priority (first represented in 1996), vehicle emissions (first represented in 1997), improved representation of flared approaches in 1998 and a better SCOOT recovery after tram priority episodes also in 1998. Version 4.4 (2002) included enhanced bus priority, enhanced gating, and a connection of already-installed emissions models to the SCOOT optimiser; allowing SCOOT to use signal timings to reduce vehicle emissions.

There is now increasing pressure to use traffic signal control to reduce

bus delays, cycle delays, pedestrian delays, pollution,
queue lengths and the number of vehicles on each road link.

There is also now increasing pressure for signal engineers to design signal systems which give more roadspace to active travel (cycling and walking) so as to

(i) encourage active travel in towns in the future and also
(ii) enhance the actual and perceived safety of active travel.

Bearing in mind these pressures, in this paper we suggest a new basis for the design of urban traffic signal control systems. We suggest that the base aim of an urban traffic control system should in future be to

maximise network capacity
on the assumption that (2)
route choices are not fixed.

Compared to our understanding of the initial SCOOT objective stated in (1) above, there are two changes in (2) here: we are proposing to

maximise network capacity instead of *minimise [(total delay) + w×(total number of stops)];*

and secondly we assume that

route-choices are not fixed instead of *route-choices are fixed*.

Both changes are important.

Typically under natural conditions there appear to be many capacity-maximising signal control policies. In this paper we consider just one rather simple capacity maximising control policy which we call:

upstream-gating merge-control.

This signal control strategy seeks to maximise network capacity while protecting specified downstream sub-networks from long queues and congestion, taking natural account of users' route-choice. There are we believe many opportunities for developing the basic ideas presented here and for applying these ideas in real life.

Under reasonable conditions the upstream-gating merge-control policy specified here maximises the capacity of the network in figure 2 subject to the two constraints:

Constraint$_1$: the traffic flow must be a user equilibrium and *Constraint*$_2$: the queue on link 3 = 0.

1.4. Traffic signal control and route choice

Suppose now that green-times are determined by using a certain responsive traffic signal control policy *P*.

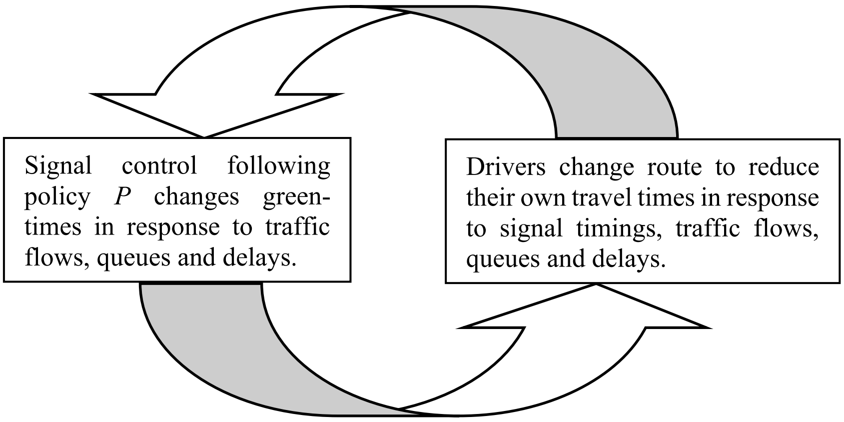

Figure 1. The day to day interaction between
(i) green-times following a responsive traffic control policy *P* and (ii) drivers' route-choice changes.

Given the responsive traffic signal control policy *P,* figure 1 is a simple representation of the interaction between green-times following the traffic signal control policy *P* and drivers' route choices. Figure 1 illustrates control policy *P* readjusting some signal timings in light of traffic flows, delays and queues at the same time as some drivers continually swap routes toward better routes. *To be specific we here think of the dynamical system in figure 1 as a day-to-day dynamical system.* But this figure may also be used to represent within-day dynamics.

The figure is based on the observation that often car drivers perform repeated journeys and so are able to learn from previous experiences. In this case it is natural to assume that over time some drivers are likely to change route if they think a different route may be better. Of course, typically, changed routes lead to changed network queues and flows which in turn will often lead the signal control policy *P* to change the signal timings to better suit the changed flows, delays and queues.

Two of the signal control policies we consider in this paper seek to take reasonable account of the dynamical system represented in Figure 1. These control policies seek to ensure that the figure 1 dynamical system leads to reasonably successful outcomes. For example we would certainly like to ensure that as the dynamical system in figure 1 evolves the queues on the network do not grow larger and larger and larger . . . .

There are many other properties we would like the dynamical system in figure 1 to have: including for example *stability.* It may be shown that with certain variations of the capacity maximising policies described in this paper the figure 1 dynamical system does indeed under certain assumptions becomes stable. See Smith and Van Vuren

(1993) for an early study of the stability of the figure 1 dynamical system. Mounce (2006, 2009) and Mounce and Carey (2011) study re-routeing in dynamic models including responsive control.

It is noteworthy that most substantial proposed (non-control) infrastructure changes in an urban or inter-urban road network are evaluated on the assumption that route choices (or mode choices) may change in light of the changes being considered. This is to take some account of travellers likely reactions to the proposed changes. In this paper we follow this with signal control policies for two distinct reasons:

(a) travellers do often have the option of changing their route and so this should be recognised by the signal control strategy design, and

(b) we wish to *positively encourage* route choice changes which benefit network performance (however defined).

1.5. Traffic signal control and route choice: capacity-maximising control policies

The basic idea behind capacity maximisation. For a given network, the signal control policy *P* is *capacity-maximising* if for any feasible demand the dynamical system illustrated in figure 1 has an equilibrium or steady state solution when the signal control policy *P* is adopted.

Thus signal control policy *P* is *capacity maximising* if, for each feasible demand, there is a (flow, green-time) pair meeting that demand such that

the flows do not change and the green-times do not change if the figure 1 dynamical system is followed.

In this case, in figure 1, flows do not change so all drivers are on their best routes and the green-times do not change so all green-times satisfy policy *P.*

Another way of putting this is as follows:
For any feasible steady origin-destination demand D; there is a (flow, green-time) pair which
(a) is consistent with demand D,
(b) is consistent with the control policy P (and so green-times stay constant in figure 1) and
(c) has all flow on cheapest routes (and so there is no incentive to change route; flows stay constant in figure 1).

Local traffic control policies which maximise network capacity or reduce queueing, under certain conditions, have been suggested by Smith (1979a, b, c, 1980, 1984, 1987) and Smith et al. (1987, 2019a, b, 2022, 2023). Maximising network capacity appears to be a much more achievable objective than total travel time minimisation. We suggest that

(i) many stakeholders would welcome seeking to maximise network capacity instead of seeking to minimise delay + [w×stops] and
(ii) many stakeholders, including car drivers, would also welcome attempts to reduce queues.

It appears that maximising network capacity may well at least in certain cases go hand in hand with reductions in queueing.

It may be shown that with certain variations of the capacity maximising policies described in this paper the figure 1 dynamical system becomes stable. See Smith and Van Vuren (1993) for an early study of stability of this dynamical system. Mounce (2006, 2009) and Mounce and Carey (2011) study re-routeing in dynamic models including responsive control. Recently Satsukawa et al (2025) show that hysteresis can occur during the figure 1 adjustment process when demand varies.

Wardrop (1952) formulates the notion of traffic equilibrium and Webster (1958) is credited with initiating the optimisation of traffic signal timings at a single junction.

1.6. Contributions of this paper

The main contributions of this paper concern the control of two simple example networks. There are three main contributions:

(1) an illustration of the capacity maximising properties of various capacity-maximising control policies,

(2) an illustration of impacts of different capacity maximising control policies on bus delays when there is no bus priority and when there is bus priority and

(3) developments of capacity maximising control policies to include upstream gating, aiming to control the location and extent of downstream queues.

2. Simple examples to motivate "capacity-maximising" traffic signal control policies (with vertical queueing) and to illustrate biased forms of the P_0 control policy

In this section we use the network in Figure 2 to illustrate and motivate the combined assignment and control approach taken in the whole paper, including the concept of "capacity-maximising" traffic signal control policies.

In this section 2 we will suppose vertical queuing. This is so as to make calculations simple in this section. We also make other assumptions here for clarity.

2.1. Equilibrium route choice

The route-choice principle adopted here and throughout this paper follows the Wardrop (1952) notion: for each origin-destination pair, more costly routes are not used. We suppose that cost equals travcl time and is measured in seconds.

The equilibrium model described in this paper utilises green-times, route-flows, link flows, bottleneck delays and queues. This model was introduced in Smith et al. (2019a, b), and follows the Thompson and Payne (1975) vertical queueing model to represent queues and queueing delays.

2.2. A simple network

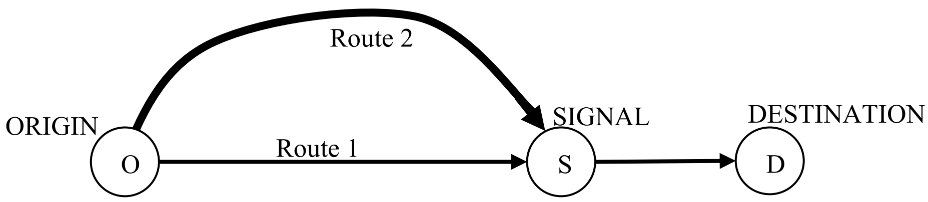

Figure 2. A simple two route network with one signal at node *S*. Route 2 is longer and wider than route 1. Route 1 and route 2 saturation flows at the signal are s_1 and s_2 vehicles / second where, importantly, $s_1 < s_2$.

2.2.1. Notation for this simple network

For $i = 1, 2$:

s_i = saturation flow at route i approach to the signal *(vehicles/second),*
C_i = freeflow travel time along route i *(seconds),*
G_i = the proportion of time route i is green,
X_i = flow along route i *(vehicles/second),*
b_i = bottleneck delay at the route i approach to the signal *(seconds),*
Q_i = queue-volume at the route i approach to the signal *(vehicles),*
t_i = travel time along route $i = C_i + b_i$ *(seconds),* and
D = the steady total flow = $X_1 + X_2$ leaving the origin for the destination *(vehicles/second)*.

We assume that $C_1 < C_2$ and importantly that

$$s_1 < s_2 \text{ and that } s_1 < D < s_2.$$

We now consider the 8-vector $(\mathbf{G}, \mathbf{X}, \mathbf{b}, \mathbf{Q}) = (G_1, G_2, X_1, X_2, b_1, b_2, Q_1, Q_2)$.

2.3. The simple network with fixed time signals

We utilise a simplified signal control model. Initially we consider the signal at *S* operating fixed-time with

the green-time proportion awarded to route 1 = G_1 = ½ and

the green-time proportion awarded to route 2 = G_2 = ½.

Suppose that the Figure 2 network is empty at time zero but from time zero onwards a steady flow of *D* vehicles per second leaves the origin for the destination, where

$$s_1G_1 < D < s_1G_1 + s_2G_2.$$

The uncongested travel time along route 1 is less than that along route 2; so we suppose that initially all *D* vehicles per second enter route 1 and no vehicles enter route 2.

Then, as soon as the steady flow along route 1 reaches node *S* a queue will start to grow on the route 1 approach to node *S*. Making simple assumptions, including that queueing is vertical, the queue volume on route 1 just upstream of node *S* will grow at a constant rate $D - s_1G_1 = D - s_1/2$ (vehicles per second). So the queueing delay b_1 on the route 1 approach to node *S* will also grow at a constant rate $D - s_1/2$ (vehicles per second); until it bcomes quicker for some traffic to choose the uncongested but longer route 2. We suppose that then some users choose route 2. In this case, typically, as time passes both routes will be used.

Now assume that as time passes the green-times, flows, delays and queues converge to an equilibrium. One such equilibrium (**G, X, b, Q**) with both routes used is:

$$G_1 = ½,\ G_2 = ½,\ X_1 = s_1/2,\ X_2 = D - s_1/2,\ b_1 = C_2 - C_1,\ b_2 = 0,\ Q_1 = X_1b_1 = (s_1/2)b_1\ ,\ Q_2 = X_2b_2 = 0.$$

This distribution of green-times, route-flows, bottleneck delays and queues

$$(\mathbf{G, X, b, Q}) = (G_1, G_2, X_1, X_2, b_1, b_2, Q_1, Q_2) = (½, ½, s_1/2, D - s_1/2, C_2 - C_1, 0, (s_1/2)b_1\ , 0)$$

is indeed, under natural conditions, a Wardrop equilibrium with vertical queueing delays, because

$$t_1 = C_1 + b_1 = C_2 + b_2 = t_2$$

and so the two route travel times are equal. This equilibrium may be regarded as a reasonable steady state description of the network performance with these fixed signal settings and vertical queues. In this paper (i) we focus mainly on Wardrop equilibria with queueing delays such as this; and (ii) as here, we do not examine closely the transient flow adjustments leading toward or away from that equilibrium state.

The steady-state models utilised in this paper may be thought of as approximations or idealisations of the peak of a peak period, where queues are at their maximum; without modelling the detailed building up or decay of those queues.

2.4.1 A responsive control policy where there is no equilibrium consistent with the policy

Consider again the simple figure 2 network. Suppose now that $s_1 < D < s_2$, again suppose vertical queueing and consider now the *equi-delay* policy:

choose green-times so that $b_1 = b_2$.

On this network, if $s_1 < D < s_2$, there is no feasible equilibrium consistent with this *equi-delay* policy. This is because this policy (equalising delays at the signal at *S*) ensures that the travel time via route 2 is always greater than the travel time via route 1, or:

$$C_1 + b_1 < C_2 + b_1 \qquad \text{since } C_1 < C_2$$

$$= C_2 + b_2 \qquad \text{since the equi-delay policy ensures that } b_1 = b_2.$$

Assuming that flow swaps to quicker routes as time passes, any flow along route 2 will in this case swap to route 1, so at any feasible equilibrium all flow must be on route 1. Since $D > s_1$, if all flow enters route 1 the inflow to route 1 must exceed the maximum possible outflow from route 1 (= s_1 *vehicles/second*). In this case the flow is unfeasible.

Thus with this equi-delay policy there can be no feasible equilibrium on this network.

Since this policy prevents equilibrium we will here say that this policy does not maximise the capacity of this network.

2.4.2 Dynamical consequences

Again suppose that $s_1 < D < s_2$. We have shown above that in this case there is no feasible equilibrium consistent with the *equi-delay* policy. A day-to-day dynamical consequence of this is that (with the equi-delay policy) if $s_1 < D < s_2$, no reasonable day-to-day dynamical model can converge to a feasible equilibrium flow pattern: because there is no feasible equilibrium consistent with the equi-delay policy.

We may also argue as follows without mentioning "equilibrium". Suppose that the steady inflow vector $\mathbf{X}(t)$ on day t is feasible; then, since $s_1 < D$, there must be positive flow on route 2 and $\mathbf{X}_2(t) > 0$. Suppose also now that the equi-delay policy is employed, and so the signal control equalises the two bottleneck delays at S in figure 1.

In this case it follows that the travel time experienced on route 2 will exceed the travel time experienced on route 1 and so it is reasonable to assume that some travellers will swap from route 2 to route 1 on day $t + 1$. As these swaps continue day after day beyond day t; on some future day the route 1 inflow will exceed s_1 vehicles per second. If this happens on day t_u then $\mathbf{X}(t_u)$ is not feasible and so $\mathbf{X}(t)$ must become unfeasible on some day t^* where t^* may be t_u or may be less than t_u.

2.5. A responsive traffic control policy where there is an equilibrium consistent with the policy

We still assume that $s_1 < s_2$.

We now consider the $P_{\mathbf{0}}$ policy:

choose green-times so that $s_1b_1 = s_2b_2$.

For example, if there are queues Q_1, Q_2 then G_1 and G_2 might be chosen so that

$Q_1/G_1 = Q_2/G_2$ or $G_1 = Q_1/(Q_1 + Q_2)$ and $G_2 = Q_2/(Q_1 + Q_2)$.

To show that there is an equilibrium consistent with this policy consider ($\mathbf{G, X, b, Q}$) where:

$G_1 = (s_2 - D)/(s_2 - s_1)$, $G_2 = (D - s_1)/(s_2 - s_1)$,

$X_1 = s_1G_1 = s_1(s_2 - D)/(s_2 - s_1)$, $X_2 = s_2G_2 = s_2(D - s_1)/(s_2 - s_1)$,

$b_1 = s_2(C_2 - C_1)/(s_2 - s_1)$, $b_2 = s_1(C_2 - C_1)/(s_2 - s_1)$,

$Q_1 = X_1b_1 = [s_1(s_2 - D)/(s_2 - s_1)][s_2(C_2 - C_1)/(s_2 - s_1)]$, $Q_2 = X_2b_2 = [s_2(D - s_1)/(s_2 - s_1)][s_1(C_2 - C_1)/(s_2 - s_1)]$.

This particular ($\mathbf{G, X, b, Q}$) is an equilibrium because

$$C_1 + b_1 = C_2 + b_2$$

so used routes have the same travel time; and also ($\mathbf{G, X, b, Q}$) satisfies the control policy since

$$s_1b_1 = s_2b_2.$$

Thus in this case, with vertical queueing, *there is an equilibrium consistent with the policy.*

So here we will say that, with vertical queueing, *this policy maximises the capacity of this network.* Compare this statement with the corresponding statement about the equi-delay policy at the end of section 2.4.1.

2.6. Another responsive control policy ($P_{\mathbf{h}}$) where there is an equilibrium consistent with the policy

Let h_1 and h_2 be any two numbers, let $\mathbf{h} = (h_1, h_2)$ and consider the policy $P_{\mathbf{h}}$:

choose green-times so that

$$s_1b_1 + h_1 = s_2b_2 + h_2.$$

(This is a biased version of $P_{\mathbf{0}}$.)

At equilibrium:

$$C_1 + b_1 = C_2 + b_2.$$

Combining these it is easy to show that:

$$b_2 = [s_1(C_2 - C_1) - (h_2 - h_1)]/[s_2 - s_1] \text{ and } b_1 = [s_2(C_2 - C_1) - (h_2 - h_1)]/[s_2 - s_1].$$

Of course bottleneck delays must be non-negative and so *this analysis and these equations only apply when*

$$s_1(C_2 - C_1) - (h_2 - h_1) \geq 0 \text{ or } (h_2 - h_1) \leq s_1(C_2 - C_1).$$

2.7. Bus delays on route 1 with P_0 and P_h without and with bus priority

2.7.1. General traffic delay on route 1 with P_0

Following section 2.5 we suppose in this P_0 case that general traffic equilibrium delay b_1 on route 1 is given by:

$$b_1 = s_2(C_2 - C_1)/(s_2 - s_1).$$

2.7.2. General traffic delay on route 1 with P_h

Following section 2.6 we suppose in this P_h case that general traffic equilibrium delay b_1 is given by:

$$b_1 = [s_2(C_2 - C_1) - (h_2 - h_1)]/[s_2 - s_1].$$

Also

$$b_2 = [s_1(C_2 - C_1) - (h_2 - h_1)]/[s_2 - s_1].$$

Since b_2 must be non-negative, the above analysis requires that

$$h_2 - h_1 \leq s_1(C_2 - C_1).$$

In particular for $h_2 - h_1 = s_1(C_2 - C_1)$:

$b_1 = [s_2(C_2 - C_1) - s_1(C_2 - C_1)]/[s_2 - s_1] = C_2 - C_1$ (and $b_2 = [s_1(C_2 - C_1) - (h_2 - h_1)]/[s_2 - s_1] = 0$).

Thus switching from P_0 to P_h reduces both b_1 and b_2 by $s_1(C_2 - C_1)]/[s_2 - s_1]$ seconds.

2.7.3. Bus delays with P_0 and P_h without bus priority and with a very simple bus priority

Suppose now that route 1 is a bus route but without bus priority. In this case we suppose that bus delays are the same as general traffic delays on route 1. See the formulae in section 2.7.1 and 2.7.2 above.

Then, by 2.7.1 above, delay to buses will be as follows:

bus delay arising with [P_0 without bus priority] = $b_1 = [s_2/(s_2 - s_1)](C_2 - C_1)$.

We here think of this as the base case.

bus delay arising with [P_β without bus priority] = $(C_2 - C_1)$.

Further, let $\beta = (s_1C_1, s_1C_2)$ and let $m = s_1/s_2$; then by 2.7.2

bus delay arising with [P_β without bus priority] = $(C_2 - C_1)$.

Therefore (using the base case "[P_0 without bus priority]") the delay reduction arising with [P_β without bus priority] as a proportion of the bus delay in the base "[P_0 without bus priority]" case is

$$\{[s_2/(s_2 - s_1)](C_2 - C_1) - (C_2 - C_1)\}/\{[s_2/(s_2 - s_1)](C_2 - C_1)\} = 1 - [s_2 - s_1]/s_2 = s_1/s_2 = m.$$

Suppose now that route 1 is a bus route *but now with bus priority*. To be both simple and precise in this case we suppose given a number R such that $0 < R < C_2 - C_1$. We now impose a very simple bus priority:

with bus priority each bus delay is reduced (compared to general traffic delay) by R seconds.

Then, subtracting R from the delays in the base case [P_0 without bus priority] (shown above) and subtracting R from the delays in the [P_β without bus priority] case (also shown above), we obtain:

bus delay arising with [P_0 with bus priority] = $[s_2/(s_2 - s_1)](C_2 - C_1) - R$ and

bus delay arising with [P_β with bus priority] = $(C_2 - C_1) - R$.

We will now compare:

1. bus delay reduction arising if [P_β without bus priority] is implemented,
2. bus delay reduction arising if [P_0 with bus priority] is implemented, and
3. bus delay reduction arising if [P_β with bus priority] is implemented.

We calculate the *proportional delay reductions* arising in each of these 3 cases using as base the bus delay arising when [P_0 without bus priority] is implemented. We have dealt with case 1, and we follow this in cases 2 and 3.

Using the above raw delay calculations, and following our case 1 calculation above, we obtain:

[P_β *without bus priority*] bus delay reduction as a proportion of the base bus delay = $m = s_1/s_2$,

[P_0 *with bus priority*] bus delay reduction as a proportion of the base bus delay = $(1-m)R/(C_2 - C_1)$, and

[P_β *with bus priority*] bus delay reduction as a proportion of the base bus delay = $m + (1-m)R/(C_2 - C_1)$.

For example if $s_2 = 2s_1$ and $R/(C_2 - C_1) = 0.1$, then

the [P_0 + *bus priority*] bus delay reduction = 0.05,
the [P_β *without bus priority*] bus delay reduction = $m = s_1/s_2 = ½$, and
the [P_β + *bus priority*] bus delay reduction = ½ + 0.05 = 0.55.

3. Illustrating the effects of different upstream-gating merge-controls on a small network

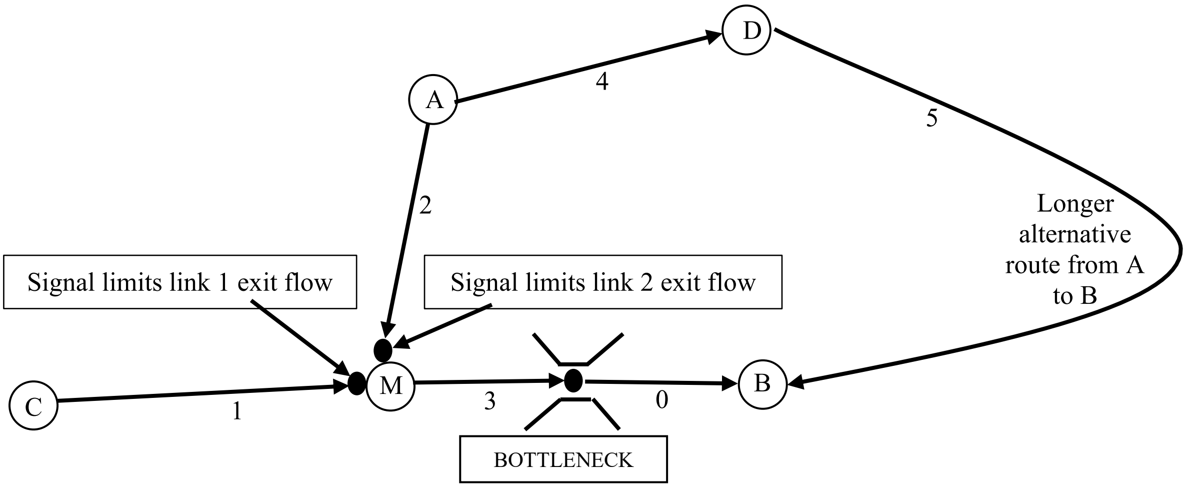

Figure 3. A network with three capacitated links (1, 2, 3) and two OD pairs [A, B] and [C, B]. Signals (black) (i) control the total exit flow from links 2 and 1 so that this total flow is within the capacity of the bottleneck link 3 (this is ensured by condition (4) below) and (ii) determine the priority to be granted to flows exiting from links 2 and 1.

We consider the network in Figure 3 where two OD pairs [A, B] and [C, B] are joined by three routes as follows:

Route 1 joins OD pair [C, B], and has links 1, 3 and 0,
Route 2 joins OD pair [A, B], and has links 2, 3 and 0, and
Route 3 (the bypass) joins the OD pair [A, B], and has links 4 and 5.

The network in figure 3 has a single main bottleneck, which is located at the exit of link 3, and a merging node for the two OD pairs, (node *M*) with two signals. These two signals can therefore be used

(i) as upstream gating control points to limit total inflows to link 3 and also
(ii) to control the merging rates of the flows along links 2 and 1 headed toward bottleneck link 3.

The demand from C to B does not have a route alternative to the route via links 1, 3 and 0 passing through the bottleneck, but the demand from A to B has the option of using a longer bypass route avoiding the bottleneck. Hence this toy network is carefully designed to analyse the combined impact of different upstream-gating and merge-control policies, accounting for the impact of route choice as a response to changing travel times due to different upstream-gating and merge-control policies. This network is also a very common part of very many real-life networks.

3.1. Notation

The main constants and variables are as follows:

s_i = saturation flow at the exit of link *i (vehicles/second, i = 1, 2, 3),*
c_i = freeflow travel time along link *i (seconds, i = 0, 1, 2, 3, 4, 5),*
v_i = flow out of link *i (vehicles/second),*
b_i = vertical queueing delay at the exit of link *i (seconds, i = 1, 2, 3),*
Q_i = queue-volume at the exit of link *i (vehicles, i = 1, 2, 3),*
T_{AB} = the (fixed or inelastic) AB flow rate (*vehicles/second*),
T_{CB} = the (fixed or inelastic) CB flow rate (*vehicles/second*),
g_1 = the proportion of time the link *1* exit is green,
g_2 = the proportion of time the link *2* exit is green,
g_3 = the *fixed* proportion of time the link *3* exit is green,

u = the capacity of the link 3 exit = s_3g_3 (*vehicles/second*),
t_2 = travel time along route 2 = $c_2 + b_2 + c_3 + c_0$ *(seconds),*
t_3 = freeflow travel time along route 3 = $c_4 + c_5$ *(seconds),* and
$\Delta = c_4 + c_5 - (c_2 + c_3 + c_0) > 0$ *(seconds).*

Thus Δ seconds is the difference between the freeflow travel time via the upper route 3 (with links 4 and 5) and the freeflow travel time via the lower route 2 (with links 2, 3 and 0) joining OD pair (A, B).

t_2 and t_3 allow us to consider equilibrium: this is a (flow, green-time, delay) pattern which meets the demand, is within the capacity constraints and is also such that no flow element has a quicker alternative route. Here this last *equilibrium* property here takes one of two forms:

Equilibrium form 1. Routes 2 and 3 are both used and $t_2 = t_3$ and
Equilibrium form 2. Route 3 is unused and $t_2 < t_3$.

3.2. Main assumptions

One central assumption here is that

$s_3g_3 = u < s_2$, $s_3g_3 = u < s_1$ and the saturation flows at the exits of links 0, 4 and 5 are large.

Thus along routes 2 and 1 the exit of link 3 must be the main bottleneck as indicated graphically in Figure 3.

Definition 1, of a feasible demand.

The demand (T_{AB}, T_{CB}) will be called feasible if

$$(T_{AB}, T_{CB}) > (0, 0) \text{ and } T_{CB} < u = s_3g_3. \quad (3)$$

3.3. Upstream gating which ensures zero queue at equilibrium on a single downstream link

The flow along link 2 = $v_2 > 0$ and the flow along link 1 = $v_1 = T_{CB}$ must both pass an "upstream-gating" signal just ahead of the merge node M, and then the flows v_2 and v_1 merge at node M.

We suppose that the upstream-gating signals, upstream of the merge at M, are to be used, primarily, to maintain a zero queue at the downstream link 3 bottleneck; by matching the total inflow to link 3 (from the merge node at M) to the outflow from link 3.

Thus the signals here will primarily control the queue on link 3 to zero by matching maximum inflow to maximum outflow on link 3. We again stress here the deterministic nature of the model so queues can become positive only in the case where the inflow is larger than the outflow; so we ensure that the maximum inflow to link 3 never exceeds the outflow from link 3.

In this paper we only consider feasible demands (T_{AB}, T_{CB})*;* that is (T_{AB}, T_{CB}) satisfying (3). We also suppose throughout this paper that the green-times g_2 and g_1 at the signals upstream of node M always restrict the total flow $v_2 + v_1$ into the merge, imposing *the upstream gating condition*

$$s_2g_2 + s_1g_1 = u = s_3g_3. \quad (4)$$

Condition (4) is our main gating condition. This condition (4) prevents the queue on link 3 from growing by ensuring that the link 3 inflow never exceeds link 3 outflow. Condition (4) has this effect since

$$[v_2 \le s_2g_2, v_1 \le s_1g_1 \text{ and (4)] implies that } v_2 + v_1 \le s_2g_2 + s_1g_1 = u = s_3g_3.$$

Here the link 3 outflow bottleneck capacity $u = s_3g_3$ is fixed. Our initial or primary control of the network is given by the upstream-gating control specified by (4). Condition (4) arose within a collaboration between Smith and Simon Parrett of the City of York Council. (Parrett proposed the term "*matching greens*" for condition (4), since this condition matches g_1, g_2 and g_3.)

For any feasible demand (T_{AB}, T_{CB}) we will also consider three *additional* merge-controls; these are the merge-controls (a) – (c) listed below. Each upstream-gating green-time vector $\mathbf{g} = (g_1, g_2)$ we consider will satisfy

the main upstream-gating condition (4) and *also* one of the three additional merge-controls (*a*) – (*c*) below.

The three additional merge-controls *(a) – (c)* are as follows:

- merge-control *(a)* (the zipper): control the flow rates along links 2 and 1 so that $v_2 = v_1$,
- merge-control *(b):* control the bottleneck delays on links 2 and 1 so that $b_1 = b_2$, and
- merge-control *(c):* control the bottleneck delays on links 2 and 1 so that $b_1 = b_2 + \Delta$.

Merge-control (*z*) will denote a "general" merge-control, i.e. one of the three merge-controls (a) – (c) listed above. For each specific merge-control (*z*), the main question we seek to address in this paper is as follows:

Is there, for each feasible demand [T_{AB}, T_{CB}], always an equilibrium consistent with both (i) the upstream-gating condition (4) and (ii) the merge-control (z)?

The answer to this question depends on the merge-control (*z*).

Definition 2. Here, in this paper, the additional merge-control (z) will be called "capacity-maximising" if for each feasible demand (T_{AB}, T_{CB}) there is an equilibrium consistent with both the upstream gating condition (4) and merge-control (z). Here (T_{AB}, T_{CB}) is feasible if (T_{AB}, T_{CB}) satisfies (3).

Note: This definition is an extension of several previous "capacity-maximising" definitions such as that above in section 1.2.1. Definition 3 is here designed to take reasonable account of the "prior" upstream-gating condition (4).

3.5. Existence and non-existence of equilibrium with a merge-control policy

It is shown in Smith et al. (2019b) that under reasonable conditions, for this network:

merge-control (a) is not capacity maximising,
merge-control (b) is capacity maximising and
merge-control (c) is capacity maximising.

Thus only *(b)* and *(c), the merge-controls which utilise delays rather than flows,* are capacity-maximising.

4. A dynamical numerical example, using the network in figure 3, with vertical queueing

This section presents a numerical example illustrating the dynamical effects of zipper policy *(a),* equi-delay policy *(b),* and the biased equi-delay policy *(c).* We utilise the network in figure 3, but now in a dynamical flow and queue context rather than an equilibrium context. We also use the "occupancies" of link 1 and link 2 as our measure of disbenefit in this section;

the occupancy of a link is to be the number of vehicles on the link.

In this section 4 we still suppose that all queueing is vertical for simplicity. Further, to be precise and also simple in this section, we suppose that:

$$s_1 = s_2 = s,$$
$$t_0 = c_1 = c_2 = c_4 > 0.$$

We have already defined Δ by:

$$\Delta = c_4 + c_5 - (c_2 + c_3 + c_0).$$

We also suppose that

$$u < s \text{ or } u/s < 1,$$

so that we must have $g_1 + g_2 < 1$ if total outflows from links 1 and 2 are to be restricted to u.

To test control policies *(a), (b)* and *(c)* we consider a single steady demand (T_{CB}, T_{AB}) *but this steady demand only starts at time 0.* This demand lasts for all time. Before time zero there are no flows within or entering the network and no queues on the network.

Now gating strategy (4) is:

$$\text{choose } (g_1, g_2) \text{ so that } s_2g_2 + s_1g_1 = u = s_3g_3.$$

In our case, where $s_1 = s_2 = s$, and $u = s_3g_3 < s$, this strategy (4) becomes:

$$\text{choose } (g_1, g_2) \text{ so that } sg_2 + sg_1 = u \text{ or } g_2 + g_1 = u/s = s_3g_3/s. \quad (5)$$

Here we consider gating strategy (4) or (5) and compare the dynamic effects of adding

merge-control *(a)* (the zipper): control the flow rates v_1, v_2, out of links 1 and 2 so that $v_2 = v_1$,
merge-control *(b)* (equi-delay): control the bottleneck delays at the exits of links 2 and 1 so that $b_2 = b_1$ and
merge-control *(c)*: control the bottleneck delays at the exits of links 2 and 1 so that $b_1 = b_2 + \Delta$.

In the text below $u/s = s_3g_3/s < 1$. We also now suppose that the network is loaded over time as follows:

$$Q_2(t) = Q_1(t) = 0 \quad \text{if } t \leqslant 0;$$
$$(T_{CB}, T_{AB})(t) = (0, 0) \quad \text{if } t \leqslant 0 \text{ and} \tag{6}$$
$$(T_{CB}, T_{AB})(t) = (3u/4, 3u/4) \quad \text{if } t > 0.$$

Thus we are supposing that inflows to links 1, 2 and 4 are zero until $t = 0$. So:

since t_0 = the time in seconds taken to traverse links 1 and 2,

queues Q_1, Q_2 at the exits of links 1 and 2 are zero until $t = t_0$.

Since (i) the link 1 and link 2 inflows start at time 0 and (ii) the travel times for links 1 and 2 are both t_0 seconds, the leading edge of the link 1 and link 2 inflows starting at time zero reach the link 1 and link 2 exit stop lines simultaneously at time $t = t_0$. Thus with vertical queueing link 1 and link 2 queues both start to grow from zero at time $t = t_0$ at a rate which initially equals

(the flow rate reaching the stop line) - (the flow rate exiting the link). (7)

4.1. Merge control (a), the Zipper, with vertical queueing.

For merge control (a), the zipper, we follow the following 6 step procedure:

Step 0. Start.

Start with an empty network, with zero flows, zero queues and zero link occupancies, at time 0. For all start times $t \geq 0$, we put

$$g_1 = g_2 = u/(2s).$$

This fixed green-time vector $(u/(2s), u/(2s))$ satisfies gating conditions (4) and (5) or

$$g_1 + g_2 = u/s = s_3g_3/s.$$

Here we keep this green-time vector

$$(g_1, g_2) = (u/(2s), u/(2s))$$

a constant for all start times $t \geq 0$.

Step 1. Initial loading of T_{CB} and T_{AB}.

For all start times $t \geq 0$, load T_{CB} and T_{AB} ($(T_{CB}, T_{AB})(t) = (3u/4, 3u/4)$ if $t \geq 0$) over time starting at time 0. Initially $(T_{CB}, T_{AB})(t)$ enters links 1 and 2. The leading edges of the *CB* and *AB* flows reach the link 1 and link 2 signal stop lines simultaneously at (say) time t_0. The link 1 and link 2 occupancies both grow at the single constant rate $3u/4$ (vehicles/sec) until time t_0.

Step 2. Provisional loading of T_{CB} and T_{AB} over all future time.

For start times $t \geq t_0$, consider persisting with the the inflow vector $(3u/4, 3u/4)$ into links 1 and 2. So long as these two link inflow rates persist for $t > t_0$, there will (for $t > t_0$) be positive vertical queues $Q_1(t)$ and $Q_2(t)$ at the exits of link 1 and 2. Therefore, since we insist that $g_1 = g_2 = u/(2s)$ for all time and the link 1 and link 2 exit saturation flows are equal, the link *1* and link *2* outflows $v_1(t)$, $v_2(t)$ must, for $t > t_0$, satisfy

$$v_1(t) = sg_1 = u/2 = sg_2 = v_2(t).$$

So in this case the equi-outflow control policy *(a)* is satisfied; equalising outflows from links 1 and 2.

Since we are persisting with the the inflow vector $(3u/4, 3u/4)$ into links 1 and 2, both queues $Q_1(t)$ and $Q_2(t)$ will be zero (at $t = t_0$) and then, for times $t \geq t_0$, both queues will grow from 0 at the rates

$dQ_1(t)/dt$ = input to the stopline of link 1 – departure from link 1 = $3u/4 - sg_1 = 3u/4 - u/2 = u/4$ veh/sec;

and (8)

$dQ_2(t)/dt$ = input to the stopline of link 2 – departure from link 2 = $3u/4 - sg_2 = 3u/4 - u/2 = u/4$ veh/sec.

Thus provided the link 1 stopline and link 2 stopline inflow rates = $3u/4$,

$$Q_1(t) = (t - t_0)(u/4) \text{ and } Q_2(t) = (t - t_0)(u/4).$$

Step 3. Provisional delays and travel times.

With the provional loading in step 2, for all times $t \geq t_0$,

the travel time via route 2 $= t_0 + b_2(t) + c_3 + c_0$

$$= t_0 + Q_2(t)/(sg_2) + c_3 + c_0$$
$$= t_0 + (t - t_0)(u/4)/(sg_2) + c_3 + c_0 = t_0 + (t - t_0)/2 + c_3 + c_0.$$

In this case,

the travel time via route 2 = $t_0 + (t - t_0)/2 + c_3 + c_0 \geq$ the travel time via route 4 = $c_4 + c_5$

if

$$(t - t_0)/2 \geq c_4 + c_5 - (t_0 + c_3 + c_0) = \Delta$$

if

$$(t - t_0) \geq 2\Delta.$$

The first or least time that the bypass (viewed from A), via links 4 and 5, becomes attractive (compared to route (link 2, link 3, link 0)) must thus be $t_0 + 2\Delta$. We now put $t_1 = t_0 + 2\Delta$.

Step 4. Link inflows for $0 \leq t \leq t_1 - t_0$ *and* *revised* *link inflows for* $t > t_1 - t_0$.

Let the (link 1 inflow, link 2 inflow, link 4 inflow) vector be ($3u/4, 3u/4, 0$) for start times t such that $0 \leq t \leq t_1 - t_0$ and now

let the revised (link 1 inflow, link 2 inflow, link 4 inflow) vector be ($3u/4, u/2, u/4$) for start times $t > t_1 - t_0$.

Then for all $t > t_1$:

the queue at the exit of link 1 grows at the rate $(3u/4 - u/2) = u/4$ and

the queue at the exit of link 2 grows at the rate of $(u/2 - u/2) = 0$.

Thus with this control policy *(a),* the above fixed green-times $g_1 = g_2 = u/(2s)$ and the *revised* route inflow specifications the route 2 and route 4 flows are equilibrated at all times t. Also the vertical queue at the exit of link 1 grows without limit as time passes, at rate $u/4$ vehicles per second. The occupancy of link 1 also grows steadily as time passes at rate $u/4$ vehicles per second. Hence the increasing (a) line in figure 4 for $t > t_1 - t_0$.

4.2. Merge control (b), equi-delay, with vertical queueing.

For merge control (b) we follow a procedure similar to that in section 4.1.

4.3. Merge control (c), equi-delay with bias, with vertical queueing.

For merge control (c) we follow a procedure similar to that in section 4.1.

4.4. Results: a graphical comparison of merge control (a) (zipper), merge control (b) (equi-delay) and merge control (c) under the dynamic loading (6).

The main performance difference between (a) and (b) is that the zipper (link 1 occupancy + link 2 occupancy) graph in Figure 3 is unbounded (with slope $u/4 > 0$) as time passes whereas the equi-delay policy (link 1 occupancy + link 2 occupancy) graph is bounded as time passes.

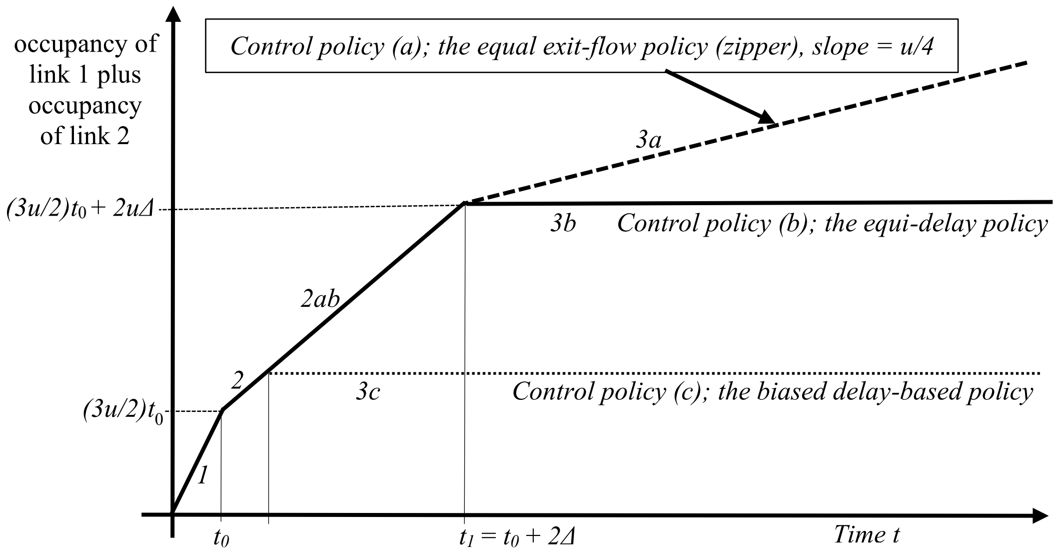

Figure 4. Trajectories of "the occupancy of link 1 plus the occupancy of link 2" when (i) the inflow (6) is loaded onto the network over time (using control policies *(a)* , *(b)* and *(c))* and (ii) all travellers choose their own best (quickest) routes. The segments *1, 2, 2ab* and *3a* have slopes $3u/2$, u, u, $u/4$.

Three trajectories are shown (corresponding to traffic control policy (a), traffic control policy (b) and traffic control policy (c)). As time passes:

if policy (a) is utilised the trajectory followed is [segment 1, segment 2, segment 2ab and segment 3a],
if policy (b) is utilised the trajectory followed is [segment 1, segment 2, segment 2ab and segment 3b] and
if policy (c) is utilised the trajectory followed is [segment 1, segment 2 and segment 3c].

All three policies utilise only local information. The main outcomes are shown in figure 4. Certain aspects of this figure may be summarised as follows:

1. Policy *(a)* utilises flows and does not maximise network capacity,
2. Policy *(b)* utilises link delays and does maximise network capacity, and
3. Policy *(c)* utilises link delays and does maximise network capacity.

Figure 4 shows that policy *(c)* yields a much higher performance than policy *(b).*

4.5. Effect on buses traversing links 1 and 3.

Buses traversing links 1 and 3 have much reduced delays with policy *(c)* compared to policy *(b).*

5. Real life results illustrating the benefits of upstream gating in Gillygate, York.

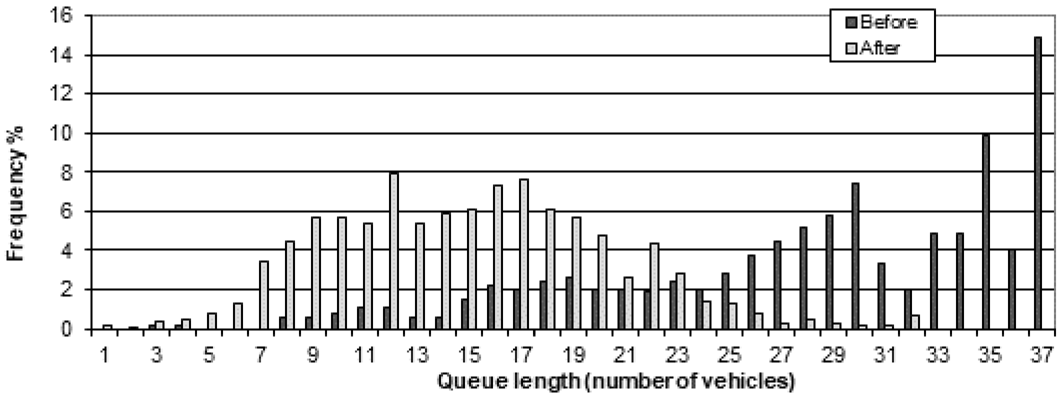

Figure 5. Queue lengths (number of vehicles) on Gillygate in York before and after the new "upstream gating" signal timing plan was introduced in 2003.

The southbound queue on Gillygate was measured at two-minutes intervals during weekdays in October 2003 (there were three measurement days prior to the implementation of the upstream-gating signal plan (these gave rise to the 'before' data in Figure 5) and eight measurement days for which the upstream-gating signal timing strategy was in operation (these days gave rise to the 'after' data in Figure 5).

Figure 5 shows (Before and After) how often the number of vehicles queueing on Gillygate between 10:00am and 3:30 pm takes on the values 1 to 37. At 37 queueing vehicles Gillygate is full. As indicated in Figure 5, Gillygate was "full" 15% percent of the time "before" the new plan and Gillygate was almost never "full" "after". Also, the most frequent number queueing "before" was 37 (full), whereas "after" it was 12. Figure 5 shows that the new "upstream gating" signal plan brought about a very significant decline in the number of queueing vehicles on Gillygate. The boxplots in Figure 6 give the mean and median of the queue length Before and After implementation of upstream gating. The dot inside each box is the mean value and the line inside the box is the median value of the Gillygate queue in this experiment.

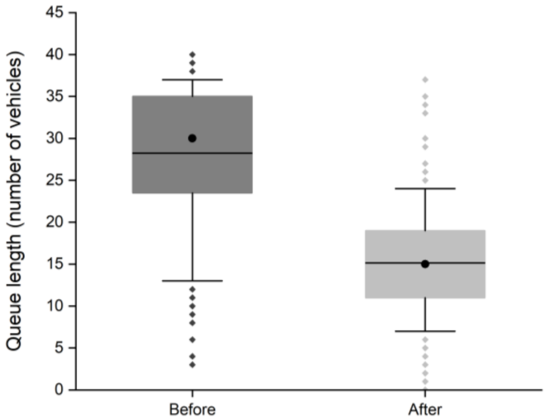

Figure 6. Queue lengths (number of vehicles) on Gillygate before and after implementation of upstream gating.

5.1. Why Gillygate?

Gillygate was chosen for this special queue-reduction treatment because Gillygate had been full of queueing traffic for decades, Gillygate has many pedestrians, and Gillygate is on several bus routes so reduced queueing here is a substantial benefit for public transport.

5.2. Part of the motivation of this paper

The above real-life Gillygate results concerned the effects of only "upstream gating". The way the total green-time was distributed between the (two) upstream approaches to Gillygate was not central to this real-life experiment.

This paper is aimed at extending upstream gating by seeking "good" ways to distribute the (already restricted) total of the upstream green-times. Section 4 presents a model within which the upstream green-times g_1, g_2 are chosen according to three different policies while also taking account of a "prior" gating restriction $g_1 + g_2 = s_3g_3/s$.

6. Conclusion.

This paper has considered responsive traffic signal control and has suggested a new approach to the design of responsive signal control policies. The paper suggests responsive controls which aim at maximising network capacity (assuming that route choices are variable), has illustrated several capacity-maximising control policies using two simple networks, and has very briefly considered their effects on buses.

The paper has shown that policy P_0 is capacity-maximising for the first network. Moreover the paper has shown that equilibrium delays arising with P_0 may be substantially reduced by using biased versions of P_0 instead of unbiased P_0.

A second simple network with a merge and a downstream bottleneck, which corresponds to a very congested part of the City of York road network – Gillygate, has also been considered. The paper has focussed on "upstream-gating" control strategies which hold traffic back at traffic signals just ahead of the merge to prevent the formation of a queue at the downstream bottleneck; and the paper has also considered different ways of further, additionally, controlling the two inflows to the merge. The paper has shown that if the added upstream merge-control uses only flows to control the two approaches to the merge, then the control is not capacity-maximising. On the other hand, the paper has shown that adding upstream merge-control which also follows P_0 is under certain conditions capacity-maximising, notwithstanding the upstream gating which reduces a downstream queue to approximately zero. This suggests that, in general, delays should probably be used to control merges and downstream queues, rather than only flows, if network capacity is to be maximised. It is shown again that biased P_0 with upstream gating is again able to very substantially reduce delays; maintaining capacity-maximisation.

The equilibrium analysis in our second network example is supported by (i) a dynamic analysis allowing for the dynamic growth of queues and (ii) real-life results of upstream-gating applied to Gillygate in York (UK) which provided motivation for this paper; these show the potential value of upstream gating in real life.

Based on the results here and other results in other papers, this paper suggests that it would be reasonable to consider upstream-gating merge-control which maximises network capacity as a new basic responsive signal control strategy for congested urban networks.

References

*The most central three references are marked *.*

Most papers listed here are available from Michael Smith or Richard Mounce.

Allsop, R.E., 1974. Some possibilities for using traffic control to influence trip distribution and route choice. In: Buckley, D.J. (Ed.), Proceedings of the 6th International Symposium on Transportation and Traffic Theory, Sydney. Elsevier, pp. 345–374.

Bretherton, D., Bowen, G., Wood, K., 2002. Effective Urban Traffic Management and Control – SCOOT version 4.4. In: Proceedings of the AET (2002).

Hunt, P.B., Robertson, D.I., Bretherton, R.D., Boyle, M.C., 1982. The SCOOT on-line traffic signal optimisation technique. Traffic Engi. Control 23, 190–192.

Mounce, R., 2006. Convergence in a continuous dynamic queueing model for traffic networks. Transp. Res. Part B 40, 779–791.

Mounce, R., 2009. Existence of equilibrium in a continuous dynamic queueing model for traffic networks with responsive signal control. In: Lam, W.H.K., Wong, S.C., Lo, H.K. (Eds.), Proceedings of 18th International Symposium on Transportation and Traffic Theory. Springer, pp. 327–344.

Mounce, R., Carey, M., 2011. Route swapping in dynamic traffic networks. Transp. Res. Part B 45 (1), 102–111.

Robertson, D.I., 1969. TRANSYT: A Traffic Network Study Tool. RRL Lab. Report LR253, Road Research Laboratory, Crowthorne, UK.

Robertson, D. I., and R. D. Bretherton. 1991. Optimizing networks of traffic signals in real time-the SCOOT method. IEEE Transactions on Vehicular Technology 40 (1): 11–15.

Satsukawa, K., Iryo, T., Yoshizawa, N., Smith, M. J., and Watling, D. 2025. Adjustment process of decentralised signal control policies with route choices: a case study with the P_0 policy. Transportmetrica B Transport Dynamics, 13(1), https://doi.org/10.1080/21680566.2025.2516183

* Smith, M.J., 1979a. A local traffic control policy which automatically maximises the overall travel capacity of an urban road network. In: Proceedings of the International Symposium on Traffic Control Systems, University of California, Berkeley, vol. 2A, pp. 11-32 (and in Traffic Engineering and Control, vol. 21, 1980, pp. 298–302).

Smith, M.J., 1979b. The existence, uniqueness and stability of traffic equilibria. Transp. Res. 13B, 295–304.

Smith, M.J., 1979c. Traffic control and route-choice; a simple example. Transp. Res. 13B, 289–294.

Smith, M.J., 1984. The stability of a dynamic model of traffic assignment– an application of a method of Lyapunov. Transportation Science 18 (3), 245–252.

Smith, M.J., Van Vuren, T., Van Vliet, D., Heydecker, B.G., 1987. The interaction between signal control policies and route choice. In: Gartner, Nathan H., Wilson, Nigel H.M. (Eds.), Proceedings of the Tenth International Symposium on Transportation and Traffic Theory. Elsevier, pp. 319–338.

Smith, M.J., 1987. Traffic control and traffic assignment in a signal-controlled network with queueing. In: Gartner, N.H., Wilson, N.H.M. (Eds.), Proceedings of the Tenth International Symposium on Transportation and Traffic Theory. Elsevier, pp. 61–68.

Smith, M.J., van Vuren, T., 1993. Traffic equilibrium with responsive traffic control. Transp. Sci. 27 (2), 118–132.

Smith, M. J., Iryo, T., Mounce, R., Rinaldi, M., Viti, F. 2019a, Traffic control which maximises network throughput: Some simple examples. Transportation Research Part C, **107**, 211-228.

Smith, M.J., Huang, W., Viti, F., Tampere, C.T.M., Lo, H.K., 2019b. Quasi-dynamic traffic assignment with spatial queueing, control and blocking back. Transportation Research Part B, **122**, 140–166.

* Smith, M.J., Iryo, T., Mounce, R, Satsukawa, K., Watling, D., 2022. Zero-queue traffic control, using green-times and prices together. Transportation Research Part C, **138**, 103630.

* Smith, M., J., Viti, F., Huang, H., Mounce, R. 2023. Upstream-gating merge-control for maximising network capacity: with an application to urban traffic management. Transportation Research Part C 155, 104287.

Thompson, W.A., Payne, H.J., 1975. Traffic assignment on a transportation network with capacity constraints and queueing. Paper presented at the 47th National ORSA/TIMS North American Meeting, 1975.

Wardrop, J.G., 1952. Some theoretical aspects of road traffic research. Proc. Inst. Civil Eng. II (1), 235–278.

Webster, F.V., 1958. Traffic Signal Settings, Department of Transport, Road Research Technical Paper No. 39, HMSO, London.