

# Optimizing Intersections

Ruth Evers\*

Stef Proost<sup>†</sup>

July 1, 2014

## Abstract

In this paper we optimize the regulation of an intersection of two routes connecting one origin-destination pair and study the effects of priority rules, traffic lights and tolls. We show that when the intersection is regulated by a priority rule the optimal policy is generally to block one of the two routes. When the intersection is regulated by traffic lights, it can only be optimal to leave both routes open when both routes are subject to congestion or if a toll is levied.

*Keywords:* Stackelberg game, road pricing, signal setting, priority rule, Braess paradox

## 1 Introduction

Despite decades of research on the optimization of intersections, the poor regulation of intersections is a matter of huge frustration amongst many drivers today. The complexity of the problem makes the optimization of intersections a tough nut to crack. Understanding the causal mechanisms that govern the optimal regulation of intersections is therefore essential in dealing with complex network problems. Not least because externalities, resulting from the behavior of drivers, can lead to results that defy intuition.

The first traffic signal control model was studied by Webster (1958). This model assumes traffic flows to be unaffected by the signal settings. This reduces the model to an isolated control problem in which signal settings are optimized for given flows on the network. The need to take into account the effects of the change in traffic light settings on the network

---

\*KU Leuven, Center for Economic Studies, Naamsestraat 69, 3000 Leuven, Belgium. Email: ruth.evers@kuleuven.be

<sup>†</sup>KU Leuven, Center for Economic Studies, Naamsestraat 69, 3000 Leuven, Belgium.

flow was first emphasized by Allsop (1974). This insight has generated two approaches in the literature that address the interaction between control and assignment: (1) the iterative procedure and (2) the global optimization approach (Cantarella et al. (1991)).

The iterative procedure iteratively solves the signal setting problem for a fixed flow pattern and the assignment problem for fixed signal settings until two successive flow patterns or signal settings converge (Allsop and Charlesworth (1977), Cantarella et al. (1991), Gartner et al. (1980), Lee and Hazelton (1996)). A dynamic approach is proposed by Hu and Mahmassani (1997) and Lo et al. (2001). With a simple example, Dickson (1981) showed that the total network cost can increase during the iterative procedure. The iterative procedure does thus not necessarily lead to the optimal solution.

When the global optimization approach is applied, some network objective function is optimized while taking into account the equilibrium route choice behaviour of the drivers (Chiou (1999), Cipriani and Fusco (2004), Fisk (1984), Marcotte (1983), Sheffi and Powell (1983), Yang and Yagar (1995)). The global optimization problem can be modelled as a bilevel programming problem in which the upper level deals with the control problem and the lower level with the user equilibrium assignment problem. The dynamic approach is studied by Abu-Lebdeh and Benekohal (2003).

Smith (1979) provided the necessary mathematical fundamentals of the traffic control and assignment problem by stating the conditions under which the problem has a unique and stable solution. In later work, he proposes and elaborates a local traffic control policy (called  $P_0$ ) that maximizes network capacity (Smith (1980, 1981)). In successive papers, Smith extends his work on the combined traffic assignment and control problem (see e.g. Smith and Van Vuren (1993)).

Relatively few papers address both signal optimization and road pricing. Clegg et al. (2001) explored the use of both instruments by specifying an algorithm that continuously moves current traffic flows, green-times and road prices within the model toward locally-optimal values. Chiou (2007) proposes a globally convergent iterative scheme designed to heuristically search for a local optimum.

Fisk (1984) was the first to model the combined assignment and control problem as a Stackelberg game. Chen and Ben-Akiva (1998) developed a dynamic model dealing with the combined assignment and control problem and formulated it as a Cournot, Stackelberg and monopoly game. Overall, there have only been a limited number of authors focussing on the game theoretical perspective of the combined assignment and control problem. For an overview, see Hollander and Prashker (2006).

This paper follows Fisk's example in modelling the combined assignment and control prob-

lem as a Stackelberg game. In this game the leader (traffic authority) moves first and bases his decision on the expected reaction of the follower (the drivers). The behaviour of the drivers can also be represented as a game, because the congestion on one road is dependent upon how many users choose to use the same road. We will assume that the followers behave according to the noncooperative principle of Nash (1951).

To obtain an optimal and tractable solution for the intersection problem, we use a simple two-road intersection, that can be controlled by either a priority rule or traffic lights. This simplified network structure can represent different types of routes and different modes of transport that are either congestible or insensitive to congestion.

The remainder of this paper is organized as follows: Section 2 focuses on the case in which the intersection is regulated by a priority rule. Section 3 presents the main results when the intersection is regulated by traffic lights. Section 4 offers a comparison between the regulation by traffic lights and by a priority rule. Section 5 illustrates the theory by means of two applications and Section 6 offers a conclusion.

## 2 Priority

In this section, the intersection is regulated by a priority rule. In Section 2.2, Route 2 is considered to have unlimited capacity, while Route 1 is subject to congestion. In Section 2.3, we consider the case where both routes have limited capacity. The focus will be on the main results and the intuition. The complete mathematical derivations are given in Appendices. Before turning to the results, we will first set out the assumptions underlying the priority model.

### 2.1 Assumptions underlying the priority model

$N$  drivers want to go from point A to point B (Figure 1). They can either take Route 1 or Route 2. The two routes intersect at point C and the intersection is regulated by a priority rule. Six assumptions are imposed on the model representing the intersection regulated by a priority rule.

**A.1** Demand: total demand is inelastic and equals  $N$ .

**A.2** Homogeneous users: all users are identical and try to minimize their expected user cost. The stationary distribution of vehicles will be a Wardrop equilibrium (Wardrop (1952)).

**A.3** Arrival rate: the arrival rate is static.

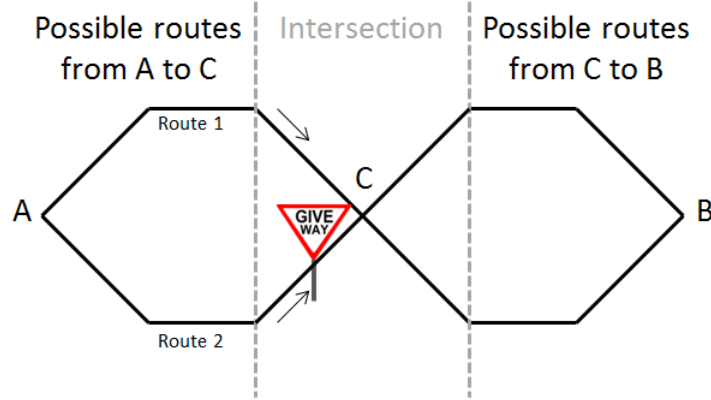


Figure 1: Outline of an intersection of two routes connecting one OD-pair (AB) regulated by a priority rule.

**A.4** Undersaturated conditions: the arrivals at the intersection on Route 2 in the time interval  $\frac{1}{X_1}$  don't exceed the number of Route 2-drivers that can cross the intersection in the time interval  $(\frac{1}{X_1} - v)$ .

**A.5** Priority: Route 1 always has priority.

**A.6** Time cost of priority:  $v$  hours before a car on Route 1 passes C, cars on Route 2 already wait until the car on Route 1 has passed.

**Proposition 1** *When the intersection of two routes connecting one OD-pair is regulated by a priority rule, the optimal policy is generally to block one of the two routes. The only exception is the case where the marginal congestion cost on the minor route is greater than half of the square of the marginal waiting cost ( $a_2 > \frac{v^2}{2}$ ). In this scenario, the optimal policy is to leave both routes open.*

## 2.2 Only one route subject to congestion

Let  $a_1$  be the increase in average cost on Route 1 when one vehicle is added, we call it the sensitivity to congestion; let  $X_i$  equal the flow on route  $i$ ; let  $\omega$  be the resource cost for a trip from A to C; let  $\phi_i$  stand for the minimal time cost from A to C of route  $i$ ; and let  $\frac{v^2 X_1}{2}$  be the expected waiting time cost<sup>1</sup> on Route 2 at the intersection. When two parallel routes connect one OD-pair, the government can use three possible policies: to block Route 1, to block Route 2, or to leave both routes open. The users in turn can react in three ways

<sup>1</sup>The waiting time cost function is derived in F.

to any chosen policy: to only take Route 1, to only use Route 2, or to use both routes. If the government decides to block Route 1, then all drivers need to take Route 2 and the total cost will equal  $(\omega + \phi_2)N$ . If, on the other hand, the government blocks Route 2, then the user equilibrium will be  $X_1 = N$  and the total cost will be  $(a_1N + \omega + \phi_1)N$ . If, however, the government decides to leave both routes open, the equilibrium reaction of the drivers will be  $X_1 = N$  if  $a_1N + \phi_1 < \phi_2 + \frac{v^2(N-1)}{2}$ ,  $X_2 = N$  if  $\phi_1 > \phi_2$  and the drivers will use both routes if  $\phi_1 < \phi_2 < a_1N + \phi_1 - \frac{v^2(N-1)}{2}$ . In this last case the Wardrop equilibrium implies  $\left(\omega + \phi_2 + \frac{v^2X_2^e}{2}\right)N$  as total cost. The user equilibrium in which both routes are used is, however, never optimal from the government's point of view. Therefore, it is always optimal to block one of the two routes.

Which route to block depends on the relative cost of both routes: a rational authority minimizing the social cost closes Route 1 if  $a_1N + \phi_1 > \phi_2$ , and Route 2 if  $a_1N + \phi_1 < \phi_2$ . Remark that if the interior equilibrium exists, it is always optimal to block the route with limited capacity. This result can be explained intuitively. If both routes are used, drivers on the minor (uncongested) route incur a waiting cost, whereas if only the minor (uncongested) route were to be used, no waiting cost would be incurred and, compared to the interior solution, no other additional costs are incurred. If on the other hand only the congested route were to be used, then the total cost would be higher than for the interior solution due to additional congestion costs. We have shown this result for linear average cost functions, but this result holds more generally for any travel cost function in which the running time is a continuous nonlinear nondecreasing function of the flow. Indeed, the main driver of the result is the additional waiting time externality that is imposed when using both Route 1 and Route 2. If there is an interior solution where both user costs are equal, it is always optimal to have all users only using Route 2 as this cost is always lower by avoiding the priority waiting costs.

In the absence of a government intervention blocking one road, the driver will often make the sub-optimal choice. Indeed, whenever  $a_1N + \phi_1 - \frac{v^2(N-1)}{2} < \phi_2 < a_1N + \phi_1$ , the user equilibrium is  $X_1 = N$ , while  $X_2 = N$  would be optimal. On the other hand, whenever  $\phi_1 < \phi_2 < a_1N + \phi_1 - \frac{v^2(N-1)}{2}$  there will be an equilibrium in which both routes are used, while it would be optimal to have all drivers on Route 2. Figure 2 illustrates this second situation. The Wardrop equilibrium is given by the intersection of the average cost-curves of Route 2 and Route 1 (point G)<sup>2</sup>. It is clear that the total cost for the interior solution equals ABCD, while the total cost would only equal ABFE if Route 1 had been blocked. This can be seen as an illustration of the Braess paradox (Braess (1968)). In the Braess paradox, adding one additional link can increase total travel cost. Braess' paradox occurs because the congestion externality is not taken into account by the drivers. Here, we also

---

<sup>2</sup>Note that for interior solutions every additional user on Route 1 imposes an extra waiting cost for the drivers on Route 2. This explains the upward sloping AC curve of Route 2 for increasing  $X_1$ .

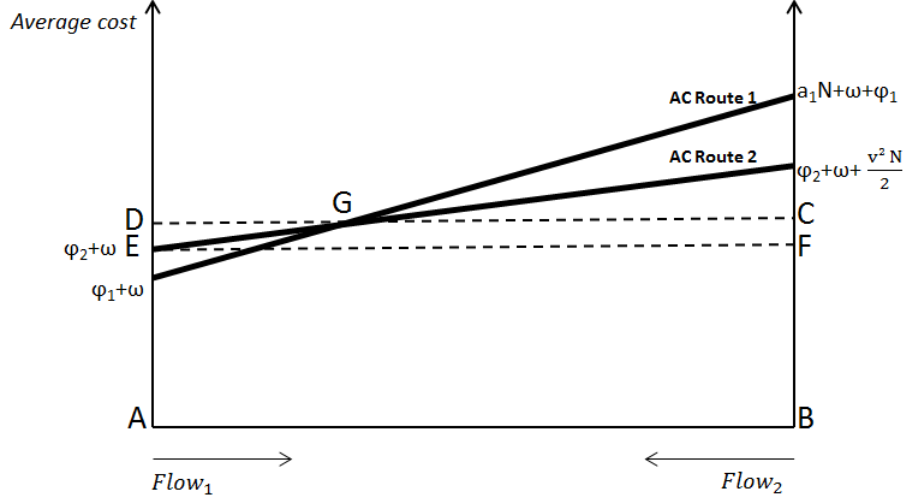


Figure 2: The solution in which all travellers use Route 2 (E) is optimal. However, without intervention, the interior solution (G) will be the user equilibrium.

add a link and, in this case, it is the external waiting cost that users on the main road impose on the users of the minor road that can increase the total travel cost.

### 2.3 Both routes subject to congestion

The suboptimality of an interior solution continues to hold for the case in which both routes are subject to congestion and  $a_2 \leq \frac{v^2}{2}$  (with  $a_2$  the sensitivity to congestion of Route 2). However, when  $a_2 \geq \frac{v^2}{2}$ , the total cost can be at lowest when both routes are used in equilibrium.

In the Stackelberg game the traffic authority (leader) moves first and bases his decision on the expected reaction of the drivers (follower). Figure 3 illustrates this sequential game and shows the different options for the government and the possible reactions of the drivers. When the government decides to close Route 1, the user equilibrium will be  $X_2 = N$ , and the total cost will be  $(a_2 N + \omega + \phi_2) N$ . When only Route 1 is accessible,  $X_1 = N$  will be the only equilibrium, and the total cost will amount to  $(a_1 N + \phi_1 + \omega) N$ . When both routes are accessible, the user equilibrium that will be in place depends on the relative value of the parameters:  $X_2 = N$  if  $a_2 N + \phi_2 < \phi_1$ ;  $X_1 = N$  if  $a_1 N + \phi_1 < \phi_2 + \frac{v^2}{2} (N - 1)$ ; and  $0 < X_1^e < N$  in all other cases.

Taking into account the reaction of the drivers, the government will block one of the two

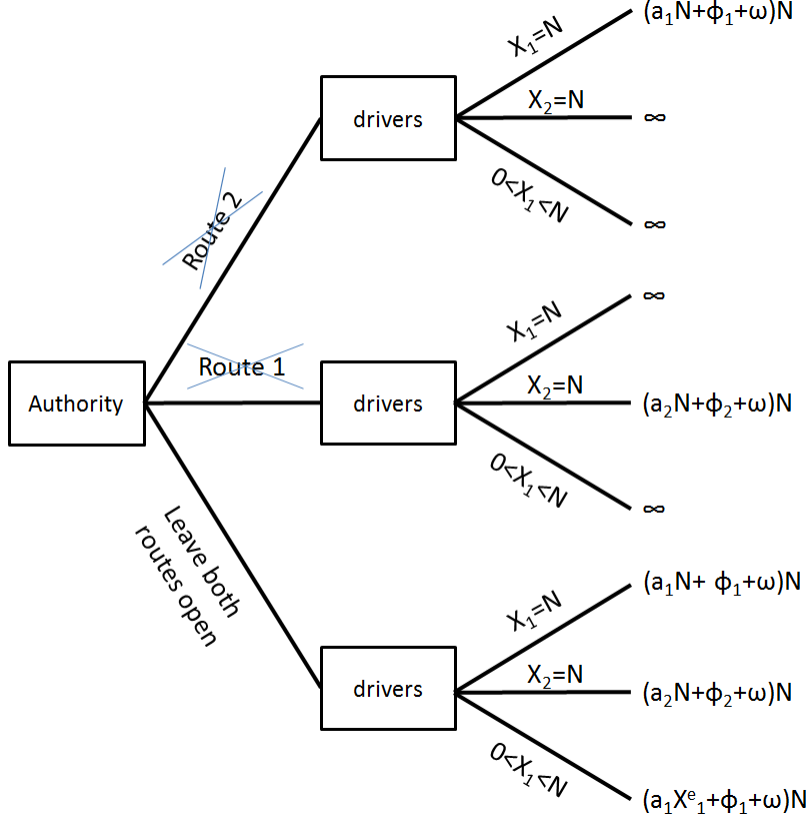


Figure 3: The total travel costs resulting from a Stackelberg game.

routes if  $a_2 < \frac{v^2}{2}$ , and leave both routes open if  $a_2 \geq \frac{v^2}{2}$ . The optimal policy thus depends on the ratio of the congestion coefficient ( $a_2$ ) to the reaction time ( $v$ ). If  $a_2 < \frac{v^2}{2}$ , the situation is similar to the case where only one route is subject to congestion and then it is always optimal to block one of the two routes. If  $a_2 \geq \frac{v^2}{2}$  and the government leaves both routes open, it can be shown that in the user equilibrium the lowest cost will be attained. In graphical terms,  $a_2 < \frac{v^2}{2}$  boils down to an upward sloping (in  $X_1$ ) average cost curve of Route 2, while  $a_2 > \frac{v^2}{2}$  boils down to a downward sloping  $AC_{route2}$ -curve. The underlying logic is that if  $a_2 \geq \frac{v^2}{2}$ , the extra congestion cost of having all travellers on Route 2 is more costly than the saving in waiting costs. If, however,  $a_2 < \frac{v^2}{2}$ , the opposite is true.

### 3 Traffic lights

In this section, the two-road intersection is regulated by traffic lights. Following the same approach as in the previous section, first only one route is considered to have limited

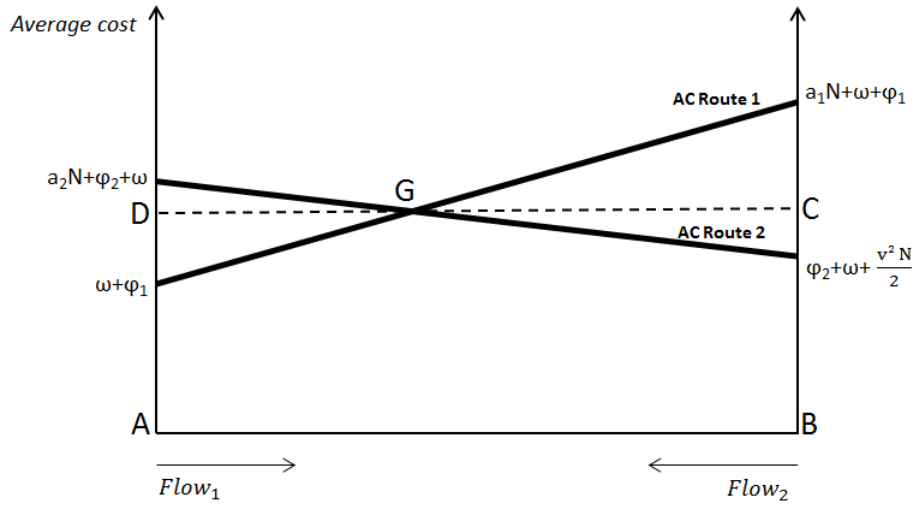


Figure 4: The interior solution (G) is optimal. No government intervention is needed to reach the optimal solution.

capacity, and subsequently both routes are considered to have limited capacity.

### 3.1 Assumptions underlying the traffic lights model

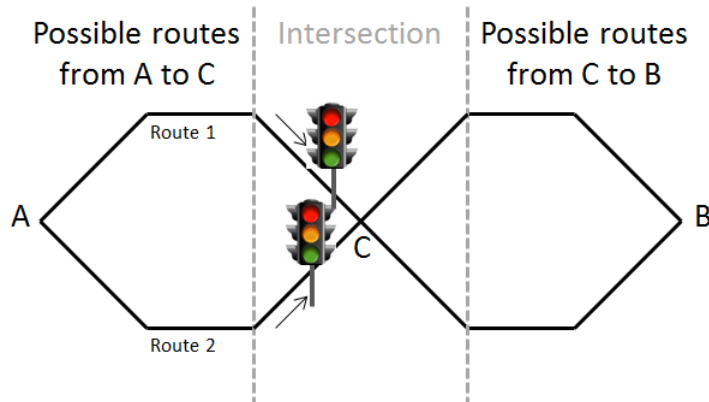


Figure 5: Outline of an intersection of two routes connecting one OD-pair (AB) regulated by traffic lights.



Five assumptions are imposed on the model representing the intersection regulated by traffic lights.

- A.1** Demand: total demand is inelastic and equals  $N$ . This assumption is relaxed in Section 3.3.2.
- A.2** Homogeneous users: all users are identical and try to minimize their expected user cost. The stationary distribution of vehicles will be a Wardrop equilibrium.
- A.3** Arrival rate: the arrival rate is static.
- A.4** Undersaturated conditions: vehicle queues are only created during red phases, and fully dissipate during green phases. This assumption is relaxed in Section 3.2.1.
- A.5** Cycle time: to simplify matters, the cycle time<sup>3</sup> ‘ $c$ ’ is held fixed. Hence, it follows that including intergreen time in the analysis is not relevant and will thus be ignored.

In this paper, the variable ‘ $r$ ’ represents the red time on Route 2 and will be the main control variable. The corresponding green time on Route 2 will thus be  $(c-r)$  and a reverse scenario holds for Route 1. As there will be alternating red times to avoid collisions at the intersection, both routes will experience an expected traffic light waiting time cost  $(T_1(c, r), T_2(c, r))$ . It is clear that the traffic light waiting cost functions are increasing in the red time and decreasing in the green time ( $\frac{\partial T_1(c, r)}{\partial r} < 0, \frac{\partial T_2(c, r)}{\partial r} > 0$ ). Furthermore, the expected traffic light waiting functions are discontinuous and jump to infinity when it is always red.

## 3.2 Only one route subject to congestion

In this section, only one route is considered to have limited capacity. In Section 3.2.1 the optimal policy is determined for the case in which the traffic authority can only control the traffic lights. In Section 3.2.2 the traffic authority can use both signal settings and a toll to minimize total costs.

### 3.2.1 Traffic lights without road pricing

**Proposition 2** *When the intersection of two routes connecting one O-D pair is regulated by traffic lights and only one of the two routes is congested, a signal setting whereby drivers choose to use both routes can never be an optimal policy.*

Table 1 shows that a rational government will never decide on an alternating signal setting when both routes are substitutes. Indeed, let  $T_i(c, r)$  be the expected waiting time cost on

---

<sup>3</sup>that is the duration of the sum of green time and red time.

Signal setting	$X_1 = N$	$X_2 = N$	$0 < X_1 < N$
$r = c$	$(a_1 N + \phi_1 + \omega) N$	$\infty$	$\infty$
$r = 0$	$\infty$	$(\omega + \phi_2) N$	$\infty$
$0 < r < c$	$(a_1 N + \omega + \phi_1 + T_1) N$	$(\omega + \phi_2 + T_2) N$	$(\omega + \phi_2 + T_2) N$

Table 1: The total travel cost for every (r, UE)-combination

route  $i$  at the intersection<sup>4</sup>. As  $T_i(c, r)$  is positive when  $0 < r < c$ , the total travel cost for an alternating signal setting will always be higher than for  $r = c$  or  $r = 0$ . Which of the two non-alternating signal settings will be optimal is dependent on the values of the parameters  $a_1, N, \phi_1$  and  $\phi_2$ . When  $a_1 N + \phi_1 < \phi_2$ ,  $r = c$  is the optimal solution and when  $a_1 N + \phi_1 \geq \phi_2$ ,  $r = 0$  will be implemented.<sup>5</sup>

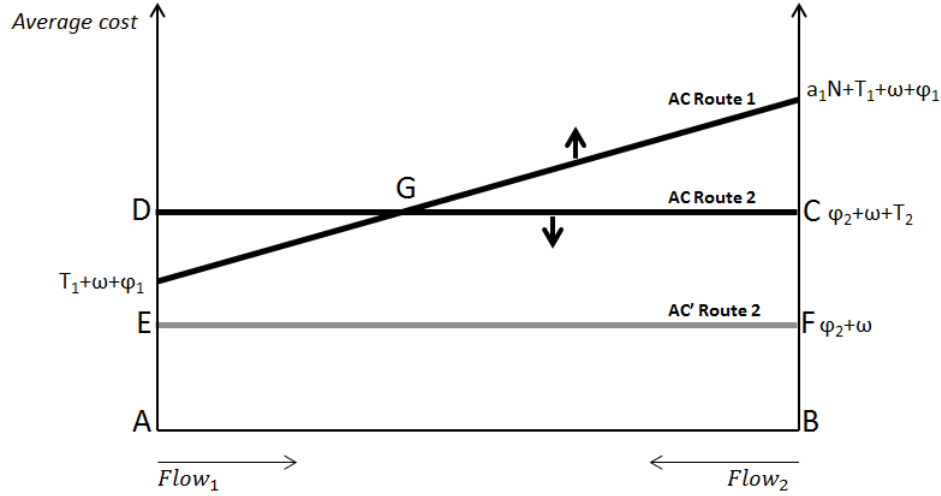


Figure 6: A corner solution constitutes the optimal solution; in this case:  $X_2 = N$ .

The suboptimality of an alternating signal ( $0 < r < c$ ) in case the intersection is regulated by traffic lights and only one route is subject to congestion, can be explained intuitively. When the user equilibrium is  $X_2 = N$  or  $X_1 = N$  (i.e.  $\phi_2 + T_2(c, r) < \phi_1 + T_1(c, r)$  or

<sup>4</sup>We will assume that the saturation flow  $s$  is very large in comparison to the arrival rate  $X_i$ , so that the traffic light waiting time due to departure delay is negligible.

<sup>5</sup>The optimality of only one route (mode) connecting one OD pair remains valid in the case where both routes have unlimited capacity. The total minimal cost then equals  $(\omega + \phi_i) N$  with  $i$  the route index that procures the lowest minimal time cost.

$a_1N + \phi_1 + T_1(c, r) < \phi_2 + T_2(c, r)$  respectively), drivers have to wait at the traffic light while there is no one crossing the intersection. The intuition behind the suboptimality of an alternating signal setting when the user equilibrium is  $0 < X_1^e < N$  is shown in Figure 6. If the duration of red light for Route 2 is reduced, then the expected average waiting time on Route 1 increases, indicated by an upward shift of the  $AC_{route1}$  curve in Figure 6. At the same time, the expected average waiting time for Route 2 decreases, corresponding to a downward shift of the  $AC_{route2}$  curve. This forces the switching point G to the left, hence less people travel on Route 1 with a simultaneous decrease in total cost. From the graph it is clear that the lowest cost (area ABFE) will be achieved when it is always green for drivers on Route 2 (grey line). It is noted that the  $AC_{route1}$  curve would lie infinitely high in this situation.

This can be vivified by the following example: consider the situation in which demand from point A to point B is relatively inelastic. Suppose A and B are connected by a tram<sup>6</sup> and a road plagued by traffic congestion. Users are indifferent between the two modes, only the user cost matters. The tram line intersects the road trajectory, and this intersection is regulated by traffic lights. In this situation, even though counterintuitive, the optimal policy would either be to close the road and only maintain the tram, or to remove the tram and only keep the road, depending on the relative costs of the two scenarios.

Proposition 2 can be generalized to average cost function in which the running time component is a continuous nonlinear nondecreasing function of the flow. Indeed, the main driver of the result is that any interior solution can be improved by either giving only green to Route 2 or giving only green to Route 1. So the result does not depend on the precise curvature of the average cost function of Route 1. Furthermore, when we relax the assumption that the waiting time due to departure delay is negligible<sup>7</sup>, the result obtained in this subsection is still valid, since any interior solution is more costly than the solution for either  $r = 0$  or  $r = c$ .

Finally, we can show that Proposition 2 can also be generalized to saturated traffic conditions. Remark first that saturated traffic conditions can only occur for alternating signal settings. Indeed, if  $r = 0$  or  $r = c$ , the traffic light can not be the restricting factor. Suppose that the demand and the relative time and resource costs are such that there exists a signal setting for which the amount of drivers arriving at the intersection on Route 1 is larger than the amount of drivers that can exit the intersection. That is, there is an  $r$  for which  $X_1^e c > s_1 r$ , with  $c$  the cycle time,  $X_1^e$  the equilibrium flow on Route 1,  $s_1$  the saturation

---

<sup>6</sup>The tram is assumed to be relatively insensitive to congestion and a substitute for the car.

<sup>7</sup>In this case, the user cost for Route 1 is  $a_1 X_1 + \phi_1 + \omega + \frac{(c-r)^2}{2c(1-\frac{X_1}{s_1})}$  and the user cost for Route 2 equals  $\phi_2 + \omega + \frac{r^2}{2c(1-\frac{X_2}{s_2})}$ .

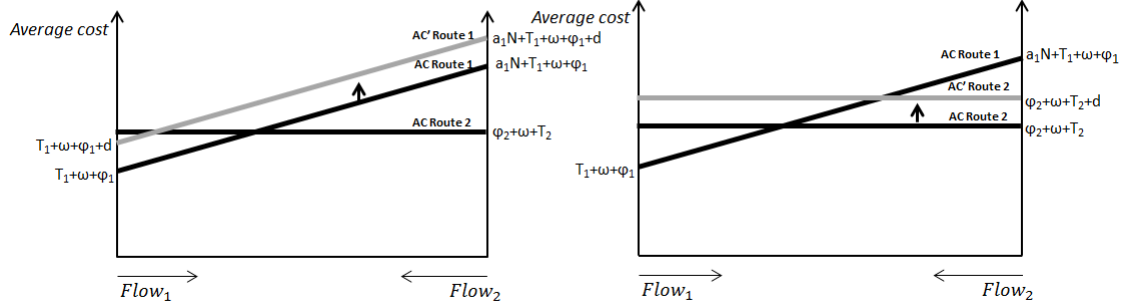


Figure 7: The total cost when traffic conditions are saturated is at least as high as when conditions are undersaturated.

flow and  $r$  the green time for Route 1 per cycle. In this case, a queue will develop on Route 1, the time cost for Route 1 will go up, and the drivers will shift from Route 1 to Route 2. If the capacity on Route 2 is not sufficient to accommodate the excess demand of Route 1, then the total demand ( $N$ ) will decrease<sup>8</sup>, which results in a loss of surplus. If Route 2 has enough spare capacity to accommodate the shifting Route 1 drivers, then we know from Yang and Yagar (1995) that in the steady state,  $X_1^e c$  equals  $s_1 r$  and the queue waiting time cost will be such that the user cost on Route 1 and Route 2 is equal. From Figure 7 (graph on the left-hand side) it is clear that the total cost in this case is as high as the total cost for the same signal setting in the undersaturated case. Figure 7 (graph on the right-hand side) furthermore shows that if Route 2 has the limiting capacity, the total cost is higher than the total cost for the same signal setting in the undersaturated case.

From the analysis of the undersaturated case, we know that the total cost for an alternating signal setting is always higher than for either  $r = 0$  or  $r = c$ . If traffic conditions can be saturated, then the optimal signal setting can lead to either saturated or undersaturated conditions. Given that we have shown that for saturated conditions, the total cost is at least as high as for the same signal setting in the undersaturated case, we can conclude that also in the saturated case it is optimal to use only one of the two routes.

### 3.2.2 Traffic lights and road pricing

In this subsection, both signal settings and a toll ( $\tau$ ) are instruments the authorities can use to minimize total costs. The toll is levied on the route subject to congestion, and the toll revenues are subtracted from the total cost. Note that with fixed demand, a toll on

<sup>8</sup>We have assumed inelastic demand. This, however, is always relatively inelastic, because if the user cost becomes infinitely large on both routes, the drivers will refrain from making the trip.

Route 1 is equivalent to a subsidy to the users of Route 2.

**Proposition 3** *When the intersection of two routes connecting one O-D pair is regulated by traffic lights and only one route is subject to congestion, then an alternating signal setting can be optimal iff a toll is possible.*

Signal setting and toll	$X_1 = N$	$X_2 = N$	$0 < X_1 < N$
$r = c$	$(a_1N + \omega + \phi_1)N$	$\infty$	$\infty$
$r = 0$	$\infty$	$(\phi_2 + \omega)N$	$\infty$
$0 < r < c, \tau > 0$	$(a_1N + \omega + \phi_1 + T_1)N$	$(\phi_2 + \omega + T_2)N$	$(a_1X_1^e + \omega + \phi_1 + T_1)N + \tau X_2^e$

Table 2: The total travel cost for every  $(\tau, r, \text{UE})$ -combination

From Table 2 it is clear that a rational authority will only implement an alternating signal setting if both routes are used in the user equilibrium. It can be shown that if there exists an optimal alternating signal setting and an optimal toll (i.e. if  $2a_1N > c$  and  $2a_1N - \frac{c}{2} + \phi_1 > \phi_2 > \phi_1 + \frac{c}{2}$ ), both routes are used. So, the optimal  $(r, \tau)$  combination can be obtained as the solution of the following minimization problem:

$$\min_{X_1, X_2, r, \tau} (a_1X_1 + \omega + \phi_1 + T_1(c, r) + \tau)X_1 + (\omega + \phi_2 + T_2(c, r))X_2 - \tau X_1 \quad (1)$$

s.t.

$$X_1 + X_2 = N \quad (2)$$

$$a_1X_1 + \omega + \phi_1 + T_1(c, r) + \tau = \omega + \phi_2 + T_2(c, r) \quad (3)$$

$$0 \leq r \leq c \quad (4)$$

$$X_1 > 0 \quad (5)$$

$$X_2 > 0 \quad (6)$$

$$\tau > 0 \quad (7)$$

Now that a toll can be levied, the external costs on Route 1 are taken into account in the user equilibrium. The optimal toll is (A):

$$\tau = \frac{\phi_2 - \phi_1 + T_2(c, r) - T_1(c, r)}{2} \quad (8)$$

Assuming undersaturated traffic conditions, i.e. queues at the intersection are only created during the red phases and dissolved during the green phases, the traffic light functions take the following form for  $0 < r < c$  (see F):

$$T_1(c, r) = \frac{(c-r)^2}{2c} \quad (9)$$

$$T_2(c, r) = \frac{r^2}{2c} \quad (10)$$

The optimal toll can then be written as follows:

$$\tau = \frac{\phi_2 - \phi_1 - \frac{c}{2} + r}{2} \quad (11)$$

As  $\partial\tau/\partial r > 0$ , the optimal toll on Route 1 is increasing in  $r$  (the green time on Route 1). The intuition is the following: if the green time increases on the congested route, people will switch from Route 2 to Route 1 to take advantage of the extra green time. However, this causes Route 1 to be even more congested. As individuals do not take into account the effect of their switching on the existing drivers on Route 1, the toll has to increase in order to reach a social optimum for given traffic light settings.

The dependence of the optimal toll on the optimal traffic light setting drives the statement that the implementation of optimal signal settings can lead to a higher acceptance of toll roads. First, in the absence of a toll, the optimal policy is to block one of the two routes. When a toll becomes available on Route 1, it can be optimal to open both routes, which increases the possibilities for the drivers. Second, consider a suboptimal signal setting without toll where the congested Route 1 receives a very low green time. When a toll can be implemented on Route 1, one can increase the green time on Route 1.

The optimal  $r$  is defined by the following equation in which  $N$  represents the total amount of vehicles,  $\tau$  the toll,  $a_1$  the sensitivity to congestion on Route 1, and  $T_i(c, r)$  the expected waiting time cost on route  $i$  at the intersection (A).

$$\left(-N + \frac{\tau}{a_1}\right) \frac{\partial T_2(c, r)}{\partial r} = \frac{\tau}{a_1} \frac{\partial T_1(c, r)}{\partial r} \quad (12)$$

Combining equations (8) and (12) allows us to identify the optimal  $r$  as the solution of the following equation:

$$(\phi_2 - 2a_1N - \phi_1 + T_2 - T_1) \frac{\partial T_2(c, r)}{\partial r} = (\phi_2 - \phi_1 + T_2 - T_1) \frac{\partial T_1(c, r)}{\partial r} \quad (13)$$

Assuming undersaturated traffic conditions and using the results from F, the optimal  $r$  as a function of the exogenous parameters can be obtained from equation (13):

$$r^* = \frac{(\phi_1 - \phi_2 + \frac{c}{2})c}{-2a_1N + c} \quad (14)$$

When  $2a_1N - c < 0$ , the optimal  $(r, \tau)$  combination is a saddle point, and consequently the minimum will be near the boundary (A). On the other hand, if  $2a_1N - c \geq 0$ , the optimal  $(r, \tau)$  combination provides the minimum attainable costs.

To identify the optimal policy, the total cost needs to be compared for all three policies. When the user equilibrium is  $X_2 = N$  or  $X_1 = N$ , tolls and traffic lights are not needed. Therefore, the total costs equal  $(\omega + \phi_2)N$  in case  $r = 0$ , and  $(a_1N + \omega + \phi_1)N$  in case

$r = c$ .

If the optimal alternating signal setting exists, i.e. if  $2a_1N > c$  and  $2a_1N - \frac{c}{2} + \phi_1 > \phi_2 > \phi_1 + \frac{c}{2}$ , then the corresponding cost is always lower than the cost for  $r = 0$  or  $r = c$  (H). If an optimal alternating signal setting does not exist, then it is optimal to implement  $r = c$  if  $\phi_2 > \phi_1 + a_1N$ , whereas it is optimal for Route 2 if  $\phi_2 \leq \phi_1 + a_1N$ .

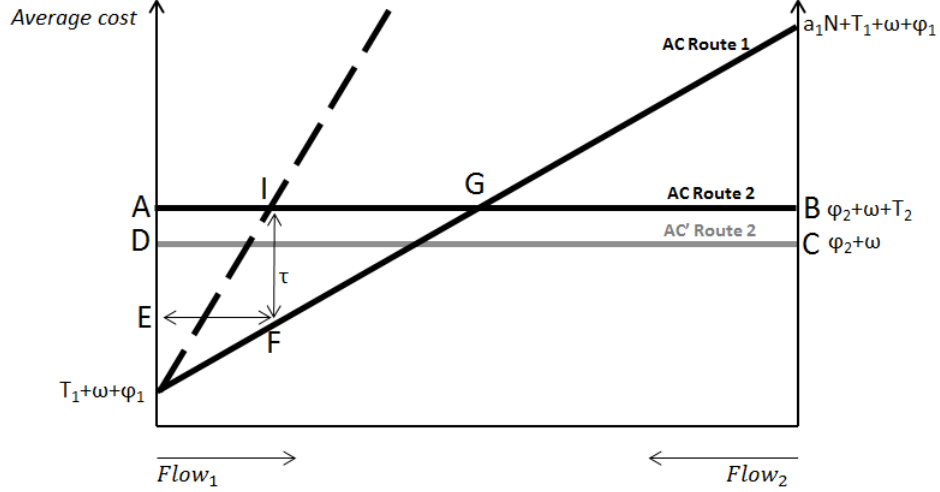


Figure 8: The interior solution (I) is optimal.

Figure 8 schematically depicts the situation in which a toll is levied on the limited capacity route in a two-road network and the intersection is regulated by traffic lights. In this graph the signal settings are considered optimal. It is clear that an interior solution exists (point I). A switch toward the corner solution  $X_2 = N$  is accompanied by a cost reduction equal to the area ABCD. This equals:

$$N \frac{r^{*2}}{2c} = N \frac{(\phi_1 - \phi_2 + \frac{c}{2})^2 c}{2(-2a_1N + c)^2} \quad (15)$$

However, the toll revenue obtained at the interior solution equals area AIFE. This is:

$$X_1 \tau(X_1) = \frac{(\phi_2 - \phi_1 + r^* - \frac{c}{2})^2}{4a_1} \quad (16)$$

In G, it is shown that area AIFE > area ABCD. Therefore, in this case, the interior optimum is better than the corner solution  $X_2 = N$ . This confirms the previous statement that if an interior optimum exists, it leads to lower costs than the corner solutions.

Figure 8 also shows that, if the optimal  $\tau$  is levied and if the traffic light settings are optimal, the user equilibrium corresponds to the social optimum. This makes sense, as the externality is internalized by the toll.

### 3.3 Both routes subject to congestion

We will first focus on the case in which demand is inelastic and then extend the analysis to include elastic demand.

#### 3.3.1 Inelastic demand

**Proposition 4** *When the intersection of two congested routes connecting one O-D pair is regulated by traffic lights, the optimal alternating signal setting is independent of the total flow and is given by  $r = \frac{a_2 c}{a_1 + a_2}$ .*

The government first decides on the signal settings and the drivers subsequently determine which route to take. Figure 9 illustrates this sequential game and shows all the feasible signal settings and the reaction of the drivers to each of these signal settings. The third branch, representing the decision of the government to implement an alternating signal setting ( $0 < r < c$ ), considers all red times between zero and  $c$ .

It can be shown that, if there exists an alternating signal setting for which the total cost is lower than for any non-alternating  $r$  (i.e., if  $a_1 X_1^e + T_1(c, r) < a_1 N$  and  $a_2 X_2^e + T_2(c, r) < a_2 N$ ), then the user equilibrium for this  $r$  is the one in which both routes are used. Hence, the optimal alternating signal setting is the solution of the following optimization problem:

$$\min_{X_1, X_2, r} (a_1 X_1 + \omega + \phi_1 + T_1(c, r)) X_1 + (a_2 X_2 + \omega + \phi_2 + T_2(c, r)) X_2 \quad (17)$$

s.t.

$$X_1 + X_2 = N \quad (18)$$

$$a_1 X_1 + \omega + \phi_1 + T_1(c, r) = a_2 X_2 + \omega + \phi_2 + T_2(c, r) \quad (19)$$

$$0 < r < c \quad (20)$$

$$X_1 > 0 \quad (21)$$

$$X_2 > 0 \quad (22)$$

The solution will be determined by:

$$a_1 \frac{\partial T_2(c, r)}{\partial r} = -a_2 \frac{\partial T_1(c, r)}{\partial r} \quad (23)$$



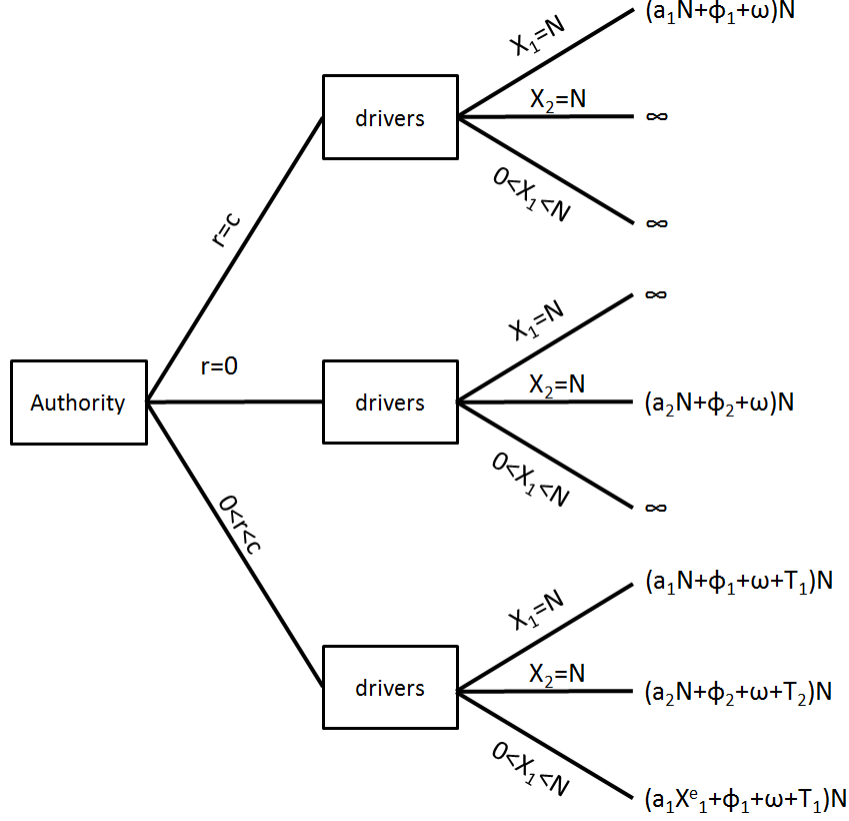


Figure 9: The total costs when the combined assignment and control problem is modelled as a Stackelberg game.

Assuming undersaturated traffic conditions, the explicit waiting time functions are given by equations (9) and (10). Inserting these expressions in equation (23), and solving for  $r$ , the proportion of red time for Route 2 equals:

$$\frac{r^*}{c} = \frac{a_2}{a_1 + a_2} \quad (24)$$

Even though the minimal time costs add to the total costs and we would therefore expect them to appear in the formula, they are not part of the optimal signal setting-formula. This can be explained by observing that the drivers themselves take the minimal time cost into account when choosing a route, while they omit the external congestion cost in their decision criterium. This omission is corrected by the optimal signal setting. The optimal red time on Route 2 increases in  $\frac{a_2}{a_2+a_1}$ . The more congestible Route 2 is compared to Route 1, the more users will take Route 1 and thus a larger cost reduction is expected from an increase in green time on Route 1.

The total cost that the optimal traffic light settings produce is given by the following equation:

$$\left( \frac{2a_1a_2(a_1+a_2)N + 2a_1(a_1+a_2)(\phi_2 - \phi_1) + a_1a_2c}{(a_1+a_2)^2} + \phi_1 + \omega \right) N \quad (25)$$

If this cost is lower than  $(a_1N + \omega + \phi_1)N$  and  $(a_2N + \omega + \phi_2)N$ , then the optimal policy is to implement  $r = \frac{a_2c}{a_1+a_2}$ . If, on the other hand, the parameters are such that  $(a_1N + \omega + \phi_1)N$  is the lowest cost, then the optimal policy would be to only give green to Route 1. Finally, if  $(a_2N + \omega + \phi_2)N$  is the lowest cost a rational authority would implement  $r = 0$ .

### 3.3.2 Elastic demand

**Proposition 5** *When the intersection of two congested routes connecting one O-D pair is regulated by traffic lights, the optimal signal setting is independent of the elasticity of demand.*

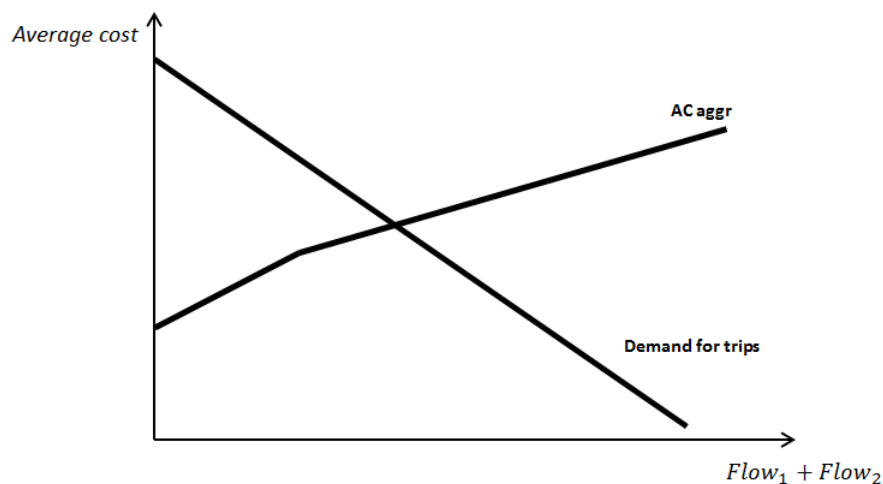


Figure 10: Equilibrium with elastic demand curve.

Regardless of the number of drivers that go from A to C, both routes will be used in the user equilibrium that comes about when the authorities implement an alternating signal setting. Taking this into account, we can construct an aggregate average cost function (C). The optimal alternating signal setting is the solution of the following maximization problem

with  $X_T$  the total flow in the network and  $(\delta - \pi X_T)$  the demand function:

$$\max_{r, X_T} \int_0^{X_T} \left( (\delta - \pi X_T) - \left( 2 \frac{a_1 a_2}{a_1 + a_2} X_T + 2 \frac{a_1 \phi_2 + a_2 \phi_1}{a_1 + a_2} + 2\omega + \frac{a_1 T_2(c, r) + a_2 T_1(c, r)}{a_1 + a_2} \right) \right) dX_T \quad (26)$$

s. t.

$$\frac{2a_1 a_2 X_T}{a_1 + a_2} + \frac{2(a_1 \phi_2 + a_2 \phi_1)}{a_1 + a_2} + 2\omega + \frac{a_1 T_2(c, r) + a_2 T_1(c, r)}{a_1 + a_2} = \delta - \pi X_T \quad (27)$$

$$X_T \geq 0 \quad (28)$$

$$c > r > 0 \quad (29)$$

$X_T$  is determined by equalizing the elastic demand function  $(\delta - \pi X_T)$  and the aggregate average cost function (27) and can be written as follows:

$$X_T = \frac{(a_1 + a_2) \delta - 2\omega(a_1 + a_2) - 2a_1 \phi_2 - 2a_2 \phi_1 - a_1 T_2(c, r) - a_2 T_1(c, r)}{(a_1 + a_2) \pi + 2a_1 a_2} \quad (30)$$

Assuming traffic conditions are undersaturated, equations (9) and (10) hold, and can be introduced in the optimization function. The optimal red time is then given by :

$$r^* = \frac{a_2 c}{a_1 + a_2} \quad (31)$$

This equation shows that the optimal red time is independent of the total flow within the network and yields the same result as in the case with inelastic demand. Therefore, it can be concluded that the optimal alternating signal settings are independent of the elasticity of demand when both routes have limited capacity.

## 4 The choice between traffic lights and a priority rule

**Proposition 6** *If only one route is subject to congestion, then for all  $r < \sqrt{v^2 c \frac{\phi_2 - \phi_1}{a_1 - \frac{v^2}{2}}}$  traffic lights are better than a priority rule. Furthermore, the higher the total number of drivers, the larger the cost advantage of traffic lights compared to a priority rule.*

If for some reason both routes have to be used and if  $r$  can be chosen such that  $r < \sqrt{v^2 c \frac{\phi_2 - \phi_1}{a_1 - \frac{v^2}{2}}}$  and  $0 < r < c$ , then traffic lights are the better choice. This can be seen as follows: if both routes have to be used when traffic lights are present, the total cost amounts to  $(\omega + \phi_2 + \frac{r^2}{2c})N$ . Comparing this with the total cost in a priority rule situation,  $(\omega + \phi_2 + \frac{v^2 X_1^e}{2})N$ , it is clear that if  $\frac{r^2}{2c} < \frac{v^2 X_1^e}{2}$  (this comes down to  $r < \sqrt{v^2 c X_1^e}$ ), traffic lights reduce the total cost. The cost savings that accompany the transfer to traffic light

regulation thus equal  $(\frac{v^2 \phi_2 - \phi_1}{2 \frac{a_1 - v^2}{2}} - \frac{r^2}{2c})N$ . If this value exceeds the additional annualized investment cost, traffic lights are optimal.

A further examination on the condition on  $r$  ( $r < \sqrt{v^2 c \frac{\phi_2 - \phi_1}{a_1 - \frac{v^2}{2}}}$ ) shows that traffic lights become more interesting when drivers are more careful (higher  $v$ ). This is explicable, a higher  $v$  increases the lost time when the intersection is regulated by a priority rule, leading to a favourable regulation of the intersection by traffic lights.

## 5 Two applications

In this section, some of the theoretical results obtained in the previous section are illustrated with an example.

The first example applies a result of Section 2.2: when the intersection of two routes connecting one OD-pair, of which only one is congestible, is regulated by a priority rule and the interior equilibrium exists, then the optimal policy is to block the congested route. The second example extends the results obtained in Section 3.2 by including local traffic within the problem setting.

### 5.1 A low-traffic city center

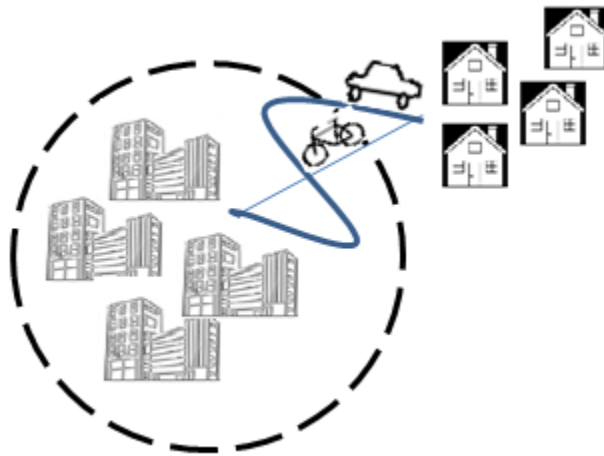


Figure 11: Commuters can take the car or the bicycle.

Consider a city where the inhabitants live on the edge of the city and work in the city center. There is a bike path, as well as a congestible road connecting work and home, and both are currently being used for commuting trips. In the city center, cyclists always have

to give way to cars. Applying the results of Section 2.2 to this situation, we can conclude that for this city, the optimal policy would be to make the city center a car-free zone. A city center in which no motorized traffic at all is allowed, is however, unrealistic. After all, shops have to be provisioned and emergency vehicles have to be able to enter the center. A simple solution, already adopted in many cities, is to allow only certain vehicles to enter the city center. This can be implemented using, for example, automatic rising bollards.

## 5.2 A bypass and a city road

A well-known situation in which one O-D pair is connected by two parallel roads is represented in Figure 12. Here, transit traffic can choose between Route 1 or Route 2 to reach point C, while local traffic can only take Route 2. Let Route 1 (the bypass) have a large capacity. Furthermore, we will assume that both local traffic and cut-through traffic contribute to the city road congestion. Suppose that the traffic lights are regulated by a federal authority whose objective is to minimize the total cost of all drivers.

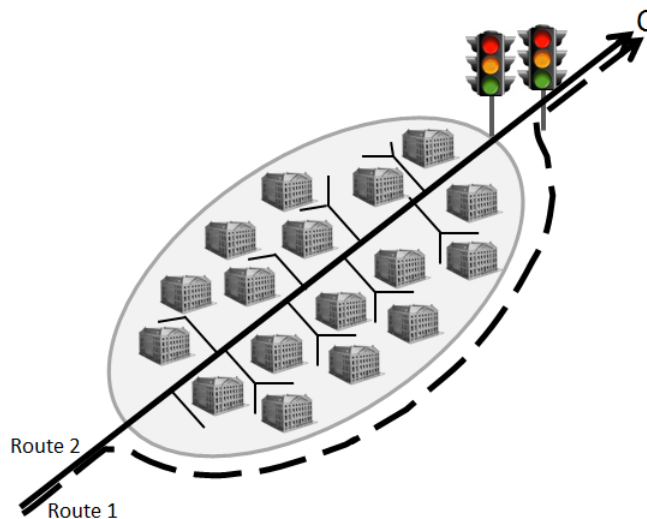


Figure 12: Transit traffic will either take the city road or the bypass depending on the signal settings.

Let

$X_b$  be the number of transit drivers taking the bypass per hour;

$X_v$  be the number of transit drivers taking the city road per hour;

$a_v$  be the congestion sensitivity of the city road;

$\phi_v$  be the minimal time cost to get to point C using the city road;  
 $\phi_b$  be the minimal time cost to get to point C using the bypass;  
 $R$  be the local traffic per hour;  
 $N$  be the total transit traffic per hour;  
 $T_i(c, r)$  be the total waiting time cost at the traffic light on route  $i$ ;  
 and  $r$  be the red time for the city road.

Signal setting	$X_b = N,$	$X_v = N$	$0 < X_b < N$
$r = c$	$\infty$	$\infty$	$\infty$
$r = 0$	$\infty$	$(\phi_v + a_v(N + R))(N + R)$	$\infty$
$0 < r < c$	$(a_v R + \phi_v + T_v)R$ $+ (\phi_b + T_b)N$	$(a_v(N + R) + \phi_v + T_v)(N + R)$	$(\phi_b + T_b)(N + R)$

Table 3: The total travel cost for every (r,UE)-combination in a bypass situation

In Table 3, the total cost is shown for every combination of policy and user equilibrium. A glance at the table shows that a rational authority would never implement  $r = c$ . When the city road always has green ( $r = 0$ ), the only Nash equilibrium is  $X_v = N$ . The total cost in this case amounts to  $(\phi_v + a_v(R + N))(R + N)$ . Furthermore, the total cost for the combination ( $X_v = N, 0 < r < c$ ) is always larger than the total cost for the combination ( $X_v = N, r = 0$ ). Finally, it can be shown that the FOC of  $(\phi_b + T_b)(N + R)$  w.r.t.  $r$  is always negative. As a consequence, the lowest cost when both routes are used occurs at the signal setting for which transit traffic is indifferent between using both routes and using only the bypass. The total cost curve when  $X_b = N$  is convex. Here, the minimum of this total cost curve equals  $r = \frac{Nc}{N+R}$ . The previous observations narrow down the candidate solutions to either  $r = 0$ ,  $r = -a_v R - \phi_v + \phi_b + \frac{c}{2}$  or  $r = \frac{Nc}{N+R}$ . In order to determine the optimal signal setting, the government has to compare the cost of the different solutions.

Suppose now that the city road has limited capacity, then how will this change the results? From Section 3.2.1, we know that if a signal setting leads to saturated conditions on the city road and both routes are used, the total cost will be the same as in the undersaturated case. Consequently, the first order condition of the total cost function will again be negative and the optimal solution will be the signal setting for which transit traffic is indifferent between using both routes and using only the bypass.

The minimal green time for which all the local drivers are below capacity ( $r = c - \frac{Rc}{s_v}$ ) will be an element of the interval for which  $X_b = N$ . If  $r = c - \frac{Rc}{s_v}$  is greater than  $\frac{Nc}{N+R}$ , then

either  $\frac{Nc}{N+R}$ ,  $r = -a_v R - \phi_v + \phi_b + \frac{c}{2}$ , or  $r = 0$ , will be the optimal solution. If on the other hand  $r = c - \frac{Rc}{s_v}$  is lower than  $\frac{Nc}{N+R}$ , then either  $r = c - \frac{Rc}{s_v}$ ,  $r = -a_v R - \phi_v + \phi_b + \frac{c}{2}$ , or  $r = 0$ , will be the optimal solution<sup>9</sup>.

It is clear that a local government, preferring minimal transit traffic in its city, would try to avoid the  $(X_v = N)$  outcome. The local government can do this by increasing  $a_v$  or  $\phi_v$ <sup>10</sup>. Increasing  $a_v$  or  $\phi_v$  raises the cost of the  $(X_v = N, r = 0)$  combination relatively more<sup>11</sup>, which decreases the likelihood of the federal government implementing  $r = 0$ .

Today, many cities already apply this strategy. Indeed, speed bumps and speed limits are put in place to increase  $\phi_v$  and local governments limit the capacity of roads to increase  $a_v$  (De Borger and Proost (2013)).

## 6 Concluding remarks

In this paper, we studied the effects of a priority rule, traffic lights, and a toll on an intersection of two routes connecting one O-D pair. We derived the intersection regulation that minimizes total travel cost, taking into account Wardrop's principles and the delay at the intersection.

We have four major results. First, if the intersection is regulated by a priority rule, the optimal policy is generally to block one of the two routes. Second, if the intersection is regulated by traffic lights, and only one route is congestible, the optimal policy is again to block one route. However, the addition of a toll allows for an optimal alternating signal setting. Third, if the intersection is regulated by traffic lights, the optimal alternating signal setting is always independent of the elasticity of demand. Finally, if only one route is subject to congestion, the superiority of a regulation by traffic lights over a priority rule becomes more likely the lower the reaction time of the drivers, and the higher the cycle time.

These results are important for three reasons. First, the counter-intuitive nature of these results confirms the importance of a good understanding of the causal mechanisms that govern the optimal regulation. Second, these insights allow to solve larger networks more efficiently as well as more effectively. More efficient, because the increased insight in the location of the optimal solution allows for a reduction in computation time. More effective, because local optima can be detected more easily. Finally, the obtained results can be applied in practice. Our results can be useful in different contexts. We primarily think about two parallel roads (e.g. a road through an urban area and a parallel road bypassing

---

<sup>9</sup>The demand function of the local drivers is assumed inelastic, so it is relatively expensive to drive back the demand. So even if for  $\frac{Nc}{N+R}$  the total cost is minimal, the loss in consumer surplus of the drivers that no longer make the trip still reduces welfare.

<sup>10</sup>We only mention the parameters the local government can influence.

<sup>11</sup> $(\frac{dT C_{r=0}}{d\phi_v} > \frac{dT C_{r=\frac{Nc}{N+R}}}{d\phi_v} \text{ and } \frac{dT C_{r=0}}{da_v} > \frac{dT C_{r=\frac{Nc}{N+R}}}{da_v})$

the city) or two parallel modes (e.g. a train and a road connecting two cities). The results in this paper can be applied to solve one particular larger network problem. In this network problem the two routes are a chain of individual components similar to the one solved in this paper. If, for every component, it is optimal to use one and the same link<sup>12</sup>, it can be concluded that it is optimal to maintain only one route. As this composition technique can only be applied to a certain type of network problems, one future research line is to extend the model to larger networks. Other future work includes the extension of the model towards multiple government levels.

## Acknowledgements

This work was supported financially by the OT/11/068-project. The authors wish to thank three anonymous referees, the Editor-in-Chief Hai Yang as well as C. Tampère, F. Viti and participants at the ITEA meeting (Northwestern 2013) for helpful suggestions on a previous draft.

---

<sup>12</sup>We can use the results from this paper to solve this problem on the component level.



## Glossary

### Parameters

$\delta$	maximum willingness-to-pay for a trip from A to B
$\omega$	resource cost per trip from A to C
$\phi_1$	minimal time cost to go from A to C using route 1
$\phi_2$	minimal time cost to go from A to C using Route 2
$\pi$	marginal willingness-to-pay for a trip from A to B
$a_1$	increase in average cost on Route 1 when one vehicle is added
$a_2$	increase in average cost on Route 2 when one vehicle is added
$c$	cycle length
$N$	total inelastic demand from A to B
$s_1$	saturation flow of Route 1
$s_2$	saturation flow of Route 2
$v$	time gap in which the users of the minor road are stationary at the intersection before the crossing of a main road user

### Control variables

$\tau$	toll fee
$r$	duration of red per cycle given to Route 2

### Variables

$X_1$	flow on Route 1
$X_2$	flow on Route 2
$X_T$	total flow in the network

**A Derivation of  $\tau^*$  and  $r^*$  and second order conditions for  $X_1 > 0$ ,  $X_2 > 0$  (only 1 route liable to congestion)**

$$\min_{X_1, X_2, r, \tau} (a_1 X_1 + \omega + \phi_1 + T_1(c, r) + \tau) X_1 + (\omega + \phi_2 + T_2(c, r)) X_2 - \tau X_1 \quad (32)$$

s.t.

$$X_1 + X_2 = N \quad (33)$$

$$a_1 X_1 + \omega + \phi_1 + T_1(c, r) + \tau = \omega + \phi_2 + T_2(c, r) \quad (34)$$

$$0 < r < c \quad (35)$$

$$X_1 > 0 \quad (36)$$

$$X_2 > 0 \quad (37)$$

$$\tau > 0 \quad (38)$$

This can be rewritten as follows:

$$\min_{r, \tau} (\omega + \phi_2 + T_2(c, r)) N - \tau \left( \frac{\phi_2 - \phi_1 + T_2(c, r) - T_1(c, r) - \tau}{a_1} \right) \quad (39)$$

$$\phi_2 - \phi_1 + T_2(c, r) - T_1(c, r) - \tau > 0 \quad (40)$$

$$a_1 N - \phi_2 + \phi_1 - T_2(c, r) + T_1(c, r) + \tau > 0 \quad (41)$$

$$0 < r < c \quad (42)$$

$$-\tau < 0 \quad (43)$$

$$L(r, \tau) = (-\omega - \phi_2 - T_2(c, r)) N + \tau \left( \frac{T_2(c, r) - T_1(c, r) + \phi_2 - \phi_1 - \tau}{a_1} \right) \quad (44)$$

$$-\lambda_1 (-T_2(c, r) + T_1(c, r) + \phi_1 - \phi_2 + \tau) \quad (45)$$

$$-\lambda_2 (-\tau + T_2(c, r) - T_1(c, r) - a_1 N + \phi_2 - \phi_1) \quad (46)$$

$$-\lambda_3 (r - c) - \lambda_4 (-r) - \lambda_5 (-\tau) \quad (47)$$

Derivation to  $\tau$

$$\frac{\partial L}{\partial \tau} = \left( \frac{T_2(c, r) - T_1(c, r) + \phi_2 - \phi_1 - \tau}{a_1} \right) - \frac{\tau}{a_1} - \lambda_1 + \lambda_2 + \lambda_5 = 0 \quad (48)$$

$$\lambda_1 = 0, \quad \phi_2 - \phi_1 + T_2(c, r) - T_1(c, r) > \tau, \quad (49)$$

$$\lambda_1 (-T_2(c, r) + T_1(c, r) + \phi_1 - \phi_2 + \tau) = 0 \quad (50)$$

$$\lambda_2 = 0, \quad a_1 N - \phi_2 + \phi_1 - T_2(c, r) + T_1(c, r) > -\tau, \quad (51)$$

$$\lambda_2(-\tau + T_2(c, r) - T_1(c, r) - a_1 N + \phi_2 - \phi_1) = 0 \quad (52)$$

$$\lambda_5 = 0, \quad -\tau < 0, \quad (53)$$

$$\lambda_5(-\tau) = 0 \quad (54)$$

For  $0 < r < c$ :

$$\tau = \frac{\phi_2 - \phi_1 + T_2(c, r) - T_1(c, r)}{2} \quad (55)$$

Derivation to r

$$\frac{\partial L}{\partial r} = -\frac{\partial T_2(c, r)}{\partial r} N + \left( \frac{\tau}{a_1} + \lambda_1 - \lambda_2 \right) \left( \frac{\partial T_2(c, r)}{\partial r} - \frac{\partial T_1(c, r)}{\partial r} \right) - \lambda_3 + \lambda_4 = 0 \quad (56)$$

$$\lambda_1 = 0, \quad -T_2(c, r) + T_1(c, r) < -\tau + \phi_2 - \phi_1, \quad (57)$$

$$\lambda_1(-T_2(c, r) + T_1(c, r) + \phi_1 - \phi_2 + \tau) = 0 \quad (58)$$

$$\lambda_2 = 0, \quad T_2(c, r) - T_1(c, r) < \tau + a_1 N - \phi_2 + \phi_1, \quad (59)$$

$$\lambda_2(-\tau + T_2(c, r) - T_1(c, r) - a_1 N + \phi_2 - \phi_1) = 0 \quad (60)$$

$$\lambda_3 = 0, \quad r < c, \quad \lambda_3(r - c) = 0 \quad (61)$$

$$\lambda_4 = 0, \quad -r < 0, \quad \lambda_4(-r) = 0 \quad (62)$$

The second order conditions are given by the following equations:

$$\frac{\partial^2 L}{\partial r^2} = -\frac{N}{c} < 0 \quad (63)$$

$$\frac{\partial^2 L}{\partial \tau^2} = -\frac{2}{a_1} < 0 \quad (64)$$

$$\frac{\partial^2 L}{\partial r \partial \tau} = \frac{1}{a_1} \quad (65)$$

It is clear that if  $2a_1 N \geq c$ , then  $\frac{2}{a_1} \frac{N}{c} - \frac{1}{a_1^2} \geq 0$  and so the optimal  $(\tau, r)$  is a minimum. However, if  $N < \frac{c}{2a_1}$  the optimal  $(\tau, r)$  is a saddle point. In this case, the lowest point will be near the corner.

## B Derivation of $r^*$ for $X_1 > 0$ , $X_2 > 0$ (both routes liable to congestion)

$$\min_{X_1, X_2, r} (a_1 X_1 + \omega + \phi_1 + T_1(c, r)) X_1 + (a_2 X_2 + \omega + \phi_2 + T_2(c, r)) X_2 \quad (66)$$

s.t.

$$X_1 + X_2 = N \quad (67)$$

$$a_1 X_1 + \omega + \phi_1 + T_1(c, r) = a_2 X_2 + \omega + \phi_2 + T_2(c, r) \quad (68)$$

$$0 < r < c \quad (69)$$

$$X_1 > 0 \quad (70)$$

$$X_2 > 0 \quad (71)$$

This can be written as follows:

$$\min_r \left( a_1 \left( \frac{a_2 N + \phi_2 - \phi_1 + T_2(c, r) - T_1(c, r)}{a_1 + a_2} \right) + \omega + \phi_1 + T_1(c, r) \right) N \quad (72)$$

s.t.

$$a_2 N + \phi_2 - \phi_1 + T_2(c, r) - T_1(c, r) > 0 \quad (73)$$

$$a_1 N - \phi_2 + \phi_1 - T_2(c, r) + T_1(c, r) > 0 \quad (74)$$

$$0 < r < c \quad (75)$$

The corresponding Lagrangian is given by:

$$L = \left( -a_1 \left( \frac{a_2 N + \phi_2 - \phi_1 + T_2(c, r) - T_1(c, r)}{a_1 + a_2} \right) - \omega - \phi_1 - T_1(c, r) \right) N \quad (76)$$

$$- \lambda_1 (-a_2 N - \phi_2 + \phi_1 - T_2(c, r) + T_1(c, r)) \quad (77)$$

$$- \lambda_2 (-a_1 N + \phi_2 - \phi_1 + T_2(c, r) - T_1(c, r)) \quad (78)$$

$$- \lambda_3 (r - c) - \lambda_4 (-r) \quad (79)$$

The FOC are as follows:

$$\frac{\partial L}{\partial r} = - \frac{a_1 N}{a_1 + a_2} \left( \frac{\partial T_2(c, r)}{\partial r} - \frac{\partial T_1(c, r)}{\partial r} \right) - \frac{\partial T_1(c, r)}{\partial r} N \quad (80)$$

$$- \lambda_1 \left( - \frac{\partial T_2(c, r)}{\partial r} + \frac{\partial T_1(c, r)}{\partial r} \right) - \lambda_2 \left( \frac{\partial T_2(c, r)}{\partial r} - \frac{\partial T_1(c, r)}{\partial r} \right) \quad (81)$$

$$- \lambda_3 + \lambda_4 = 0 \quad (82)$$

$$-a_2N - \phi_2 + \phi_1 - T_2(c, r) + T_1(c, r) < 0, \quad \lambda_1 = 0, \quad (83)$$

$$\lambda_1(-a_2N - \phi_2 + \phi_1 - T_2(c, r) + T_1(c, r)) = 0 \quad (84)$$

$$-a_1N + \phi_2 - \phi_1 + T_2(c, r) - T_1(c, r) < 0, \quad \lambda_2 = 0, \quad (85)$$

$$\lambda_2(-a_1N + \phi_2 - \phi_1 + T_2(c, r) - T_1(c, r)) = 0 \quad (86)$$

$$r < c, \quad \lambda_3 = 0, \quad \lambda_3(r - c) = 0 \quad (87)$$

$$-r < 0, \quad \lambda_4 = 0, \quad \lambda_4(-r) = 0 \quad (88)$$

If the optimal  $r$  is inserted in constraints (73) and (74), then it becomes clear that the interior solution is only feasible if:

$$2(a_1 + a_2)(a_2N + \phi_2 - \phi_1) > a_1c - a_2c \quad (89)$$

and

$$2(a_1 + a_2)(a_2N - \phi_2 + \phi_1) > a_2c - a_1c \quad (90)$$

## C Derivation of the aggregate AC

We will assume the properties of the routes connecting CB to be the same as those connecting AC.

$$AC_{1ac} = a_1 X_{1ac} + \omega + \phi_1 + T_1(c, r)$$

$$AC_{2ac} = a_2 X_{2ac} + \omega + \phi_2 + T_2(c, r)$$

$$AC_{1cb} = a_1 X_{1cb} + \omega + \phi_1$$

$$AC_{2cb} = a_2 X_{2cb} + \omega + \phi_2$$

$$\mathbf{demand} = \delta - \pi(X_1 + X_2)$$

$$X_T = X_{1cb} + X_{2cb} = X_{1ac} + X_{2ac}$$

$$X_{1ac} = \frac{AC_{ac} - \omega - \phi_1 - T_1(c, r)}{a_1} \quad (91)$$

$$X_{2ac} = \frac{AC_{ac} - \omega - \phi_2 - T_2(c, r)}{a_2} \quad (92)$$

$$a_1 a_2 X_{Tac} = (a_1 + a_2) AC_{ac} - (a_1 + a_2) \omega - a_1 (\phi_2 + T_2(c, r)) - a_2 (\phi_1 + T_1(c, r)) \quad (93)$$

$$AC_{ac} = \frac{a_1 a_2 X_{Tac}}{a_1 + a_2} + \omega + \frac{a_1}{a_1 + a_2} (\phi_2 + T_2(c, r)) + \frac{a_2}{a_1 + a_2} (\phi_1 + T_1(c, r)) \quad (94)$$

$$AC_{cb} = \frac{a_1 a_2 X_{Tcb}}{a_1 + a_2} + \omega + \frac{a_1}{a_1 + a_2} (\phi_2) + \frac{a_2}{a_1 + a_2} (\phi_1) \quad (95)$$

$$AC_{cb} + AC_{ac} = \frac{2a_1 a_2 X_T}{a_1 + a_2} + \omega + \frac{2a_1}{a_1 + a_2} (\phi_2) + \frac{2a_2}{a_1 + a_2} (\phi_1) + \frac{a_1}{a_1 + a_2} (T_2(c, r)) + \frac{a_2}{a_1 + a_2} (T_1(c, r)) \quad (96)$$

In Figure 13 the aggregated cost curve is shown. The kink in the cost curve is due to the change from using only one route, to using both routes.

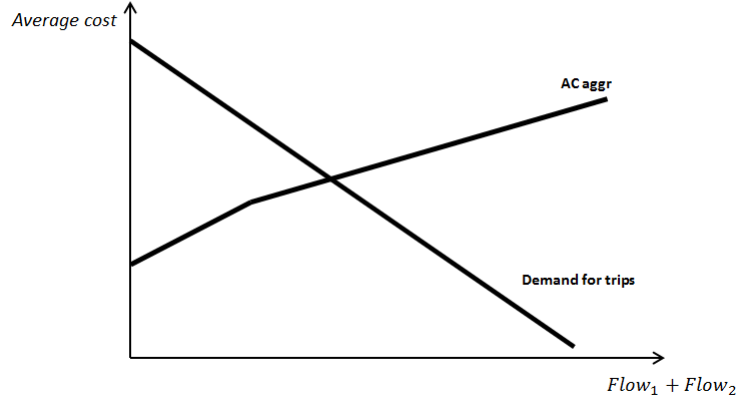


Figure 13: The aggregate cost

## D Derivation of the total cost for an intersection regulated by a priority rule (both routes liable to congestion)

$$\min_{X_1, X_2} (a_1 X_1 + \omega + \phi_1) X_1 + \left( a_2 X_2 + \omega + \phi_2 + \frac{v^2 X_1}{2} \right) X_2 \quad (97)$$

s.t.

$$X_1 + X_2 = N \quad (98)$$

$$a_1 X_1 + \omega + \phi_1 = a_2 X_2 + \omega + \phi_2 + \frac{v^2 X_1}{2} \quad (99)$$

$$X_1 > 0 \quad (100)$$

$$X_2 > 0 \quad (101)$$

The total cost at the interior solution equals:

$$(a_1 X_1^e + \omega + \phi_1) N \quad (102)$$

The total cost at the corner solution ( $X_1 = N$ ) equals:

$$(a_1 N + \omega + \phi_1) N \quad (103)$$

The total cost at the corner solution ( $X_2 = N$ ) equals:

$$(a_2 N + \omega + \phi_2) N \quad (104)$$

## E Derivation of $r^*$ when demand is elastic (both routes liable to congestion)

To find  $r^*$ , the derivative of the following equation to  $r$  is taken. After which it is checked if  $r^*$  satisfies the constraints.

The total cost equals:

$$\delta X_T^e - \frac{\pi}{2}(X_T^e)^2 - \frac{a_1 a_2}{a_1 + a_2}(X_T^e)^2 - 2\omega X_T^e - \frac{2(a_1 \phi_2 + a_2 \phi_1)}{a_1 + a_2} X_T^e - \frac{a_1 T_2(c, r) + a_2 T_1(c, r)}{a_1 + a_2} X_T^e \quad (105)$$

The derivation to  $r$  gives us the following condition:

$$\frac{\partial X_T^e}{\partial r} X_T^e \left( \frac{\pi}{2} + \frac{a_1 a_2}{a_1 + a_2} \right) = 0 \quad (106)$$

The signal setting for which  $X_T^e = 0$ , would imply that no driver would make the trip and the welfare would be equal to zero. In this case, the road could just as well be closed off.

The signal setting for which  $\frac{\partial X_T^e}{\partial r} = 0$  equals:

$$\frac{a_2 c}{a_1 + a_2} \quad (107)$$



## F Waiting time computation for one route

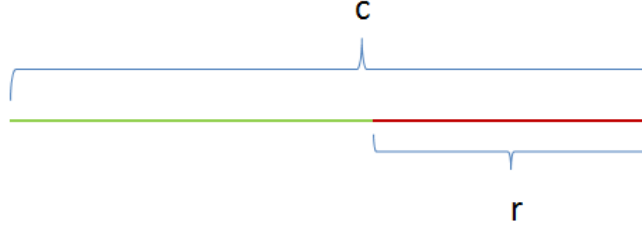


Figure 14: Composition of the cycle time

It is assumed that traffic conditions are undersaturated.

The average waiting time for drivers arriving when the traffic light is red:

$$\frac{r - \frac{d}{2} + r - \frac{3d}{2} + r - \frac{5d}{2} + \dots + \frac{3d}{2} + \frac{d}{2}}{\frac{r}{d}} = \frac{r}{2} \quad (108)$$

Average waiting time for all drivers (arriving at the traffic light during the red or green phase):

$$\frac{\frac{r}{2} \frac{r}{d}}{\frac{c}{d}} = \frac{r^2}{2c} \quad (109)$$

When the intersection is regulated by traffic lights, the waiting time for Route 2 (with r red) equals:

$$\frac{r^2}{2c} \quad (110)$$

and the waiting time for Route 1 (with c-r red) equals:

$$\frac{(c - r)^2}{2c} \quad (111)$$

In cases where the intersection is regulated by a priority rule, the red time is replaced by  $v$  ( $v$  seconds before a car on the main road passes the intersection, cars on the minor road already wait until the car on the main road has passed the intersection), and the cycle time is replaced by the interarrival time of cars on the main road ( $\frac{1}{X_1}$ ). The waiting time for cars on the minor road then equals:

$$\frac{v^2 X_1}{2} \quad (112)$$

## G Area AIFE > area ABCD

Area AIFE- area CDEF >0

$$\frac{(\phi_2 - \phi_1 - \frac{c}{2})^2}{4a_1} + \frac{(\phi_2 - \phi_1 + r^* - \frac{c}{2}) r^*}{2a_1} + \frac{r^{*2}}{4a_1} - N \frac{r^{*2}}{2c} > 0 \quad (113)$$

$$\frac{(\phi_2 - \phi_1 - \frac{c}{2})^2}{2a_1} > 0 \quad (114)$$

## H Comparison of the interior solution and the corner solutions

$TC(\text{interior}^*) < TC(X_2=N)$

$$\left( \frac{(\phi_1 - \phi_2 + \frac{c}{2})^2}{2(-2a_1N + c)} + \omega + \phi_2 \right) N < (\omega + \phi_2) N \quad (115)$$

$$\frac{(\phi_1 - \phi_2 + \frac{c}{2})^2}{2(-2a_1N + c)} < 0 \quad (116)$$

As the left hand side is negative, this is always satisfied.

$TC(\text{interior}^*) < TC(X_1=N)$

$$\left( \frac{(\phi_1 - \phi_2 + \frac{c}{2})^2}{2(-2a_1N + c)} + \omega + \phi_2 \right) N < (a_1N + \omega + \phi_1) N \quad (117)$$

$$\frac{(\phi_1 - \phi_2 + \frac{c}{2})^2}{2(-2a_1N + c)} + \phi_2 - \phi_1 < a_1N \quad (118)$$

As  $2a_1N - \frac{c}{2} + \phi_1 > \phi_2$  for an interior optimum, this is always satisfied.

## References

## References

- Abu-Lebdeh, G. and Benekohal, R. F. (2003). Design and evaluation of dynamic traffic management strategies for congested conditions. *Transportation Research Part A: Policy and Practice*, 37(2):109–127.
- Allsop, R. (1974). Some possibilities for using traffic control to influence trip distribution and route choice. In *Transportation and Traffic Theory, Proceedings*, volume 6.
- Allsop, R. and Charlesworth, J. (1977). Traffic in a signal-controlled road network: An example of different signal timings including different routeing. *Traffic Engineering and Control*, 18(Analytic).
- Braess, P.-D. D. D. (1968). Über ein paradoxon aus der verkehrsplanung. *Unternehmensforschung*, 12(1):258–268.
- Cantarella, G. E., Improta, G., and Sforza, A. (1991). Iterative procedure for equilibrium network traffic signal setting. *Transportation Research Part A: General*, 25(5):241–249.
- Chen, O. J. and Ben-Akiva, M. E. (1998). Game-theoretic formulations of interaction between dynamic traffic control and dynamic traffic assignment. *Transportation Research Record: Journal of the Transportation Research Board*, 1617(1):179–188.
- Chiou, S. (2007). A globally convergent iterative scheme for toll design network with signal settings. *Applied mathematics and computation*, 187(2):1086–1099.
- Chiou, S.-W. (1999). Optimization of area traffic control for equilibrium network flows. *Transportation Science*, 33(3):279.
- Cipriani, E. and Fusco, G. (2004). Combined signal setting design and traffic assignment problem. *European Journal of Operational Research*, 155(3):569 – 583.
- Clegg, J., Smith, M., Xiang, Y., and Yarrow, R. (2001). Bilevel programming applied to optimising urban transportation. *Transportation Research Part B: Methodological*, 35(1):41–70.
- De Borger, B. and Proost, S. (2013). Traffic externalities in cities: The economics of speed bumps, low emission zones and city bypasses. *Journal of Urban Economics*, 76:53–70.
- Dickson, T. J. (1981). A note on traffic assignment and signal timings in a signal-controlled road network. *Transportation Research Part B: Methodological*, 15(4):267–271.
- Fisk, C. (1984). Game theory and transportation systems modelling. *Transportation Research Part B: Methodological*, 18:301 – 313.

- Gartner, N. H., Gershwin, S. B., Little, J. D., and Ross, P. (1980). Pilot study of computer-based urban traffic management. *Transportation Research Part B: Methodological*, 14(1):203–217.
- Hollander, Y. and Prashker, J. N. (2006). The applicability of non-cooperative game theory in transport analysis. *Transportation*, 33(5):481–496.
- Hu, T.-Y. and Mahmassani, H. S. (1997). Day-to-day evolution of network flows under real-time information and reactive signal control. *Transportation Research Part C: Emerging Technologies*, 5(1):51–69.
- Lee, S. and Hazelton, M. (1996). Stochastic optimisation of combined traffic assignment and signal control junction modelling. In *Transportation and traffic theory. Proceedings of the 13th international symposium on transportation and traffic theory, Lyon, France*.
- Lo, H. K., Chang, E., and Chan, Y. C. (2001). Dynamic network traffic control. *Transportation Research Part A: Policy and Practice*, 35(8):721–744.
- Marcotte, P. (1983). Network optimization with continuous control parameters. *Transportation Science*, 17(2):181–197.
- Nash, J. (1951). Non-cooperative games. *The Annals of Mathematics*, 54(2):286–295.
- Sheffi, Y. and Powell, W. B. (1983). Optimal signal settings over transportation networks. *Journal of Transportation Engineering*, 109(6):824–839.
- Smith, M. (1979). The existence, uniqueness and stability of traffic equilibria. *Transportation Research Part B: Methodological*, 13(4):295–304.
- Smith, M. (1980). A local traffic control policy which automatically maximises the overall travel capacity of an urban road network. *Traffic Engineering and Control*, 21(HS-030 129).
- Smith, M. (1981). Properties of a traffic control policy which ensure the existence of a traffic equilibrium consistent with the policy. *Transportation Research Part B: Methodological*, 15(6):453–462.
- Smith, M. J. and Van Vuren, T. (1993). Traffic equilibrium with responsive traffic control. *Transportation Science*, 27(2):118.
- Wardrop, J. (1952). Some theoretical aspects of road traffic research. In *Proceedings of Institution of Civil Engineers*, volume 1, pages 325–378.
- Webster, F. (1958). Traffic signal settings. Technical report, Road Research technical paper no. 39.

Yang, H. and Yagar, S. (1995). Traffic assignment and signal control in saturated road networks. *Transportation Research Part A: Policy and Practice*, 29(2):125 – 139.