Metering-based priority with departure time choice

Raphaël Lamotte

André de Palma

Nikolas Geroliminis

Urban Transport Systems Laboratory (LUTS)

April 2019

STRC | 19th Swiss Transport Research Conference
Monte Verità / Ascona, May 15 – 17, 2019

Urban Transport Systems Laboratory (LUTS)

Metering-based priority with departure time choice

Raphaël Lamotte, Nikolas Geroliminis
Urban Transport Systems Laboratory
EPFL
Station 18 CH-1015 Lausanne
phone: +41-21-693 24 79
fax: +41-21-693 24 69
{raphael.lamotte,nikolas.geroliminis}@epfl.ch

André de Palma
ENS Cachan
CREST
University Paris-Saclay
phone: +41-21-693 24 79
fax: +41-21-693 24 69
andre.de-palma@ens-paris-saclay.fr

April 2019

Abstract

This paper explores the potential of priority schemes to alleviate congestion at a bottleneck. While advocates of reserved lanes tend to focus only on the consequences of a potential mode shift, this paper shows that prioritizing efficient modes can yield other positive effects, such that priority schemes can be successful even with a given mix of vehicle types. To that end, this paper revisits existing results known for homogeneous fleet and shows how heterogeneity in vehicle types makes priority schemes even more relevant, especially when accounting for potential adjustments in departure time. This works further examines the consequences of an endogenous mix of vehicles types and of heterogeneity in desired arrival times.

Keywords

priority, departure time, carpool, metering, heterogeneity, bottleneck

1 Introduction

Faced with the tremendous cost of congestion and strong socio-political doubts regarding congestion pricing, policymakers have turned towards alternative congestion-alleviating measures, and in particular towards those High-Occupancy Vehicle (HOV) lanes. These lanes aim at increasing the proportion of carpoolers and public transit users by providing them with reserved lanes (*spatial segregation*). Such schemes have been studied both in idealized set-ups (Dahlgren, 1998, Konishi and Mun, 2010) and with empirically estimated models (Small *et al.,* 2006). These works show that reserved lanes can improve welfare and even be Pareto-improving, but only in a restricted range of traffic conditions and potential HOV demand. These conditions are not always met in practice, which may lead to the so-called "empty lane syndrome".

In parallel, another line of research has focused on *temporal* segregation, in the context of Vickrey's departure time choice problem (Vickrey, 1969). The laissez-faire equilibrium is particularly wasteful there because of "levelling down": users have to fight (queue) for the best alternatives (passage times), which makes them less attractive. Temporal segregation reduces this queuing by reserving some highly desirable passage times for some priority users, such that they do not have to compete with all the others. If demand is homogeneous and inelastic, temporal segregation is Pareto-improving regardless of the proportion that is prioritized (Fosgerau, 2011).

So far, the literature on temporal segregation has never aimed at prioritizing efficient modes, like HOVs. In fact, it is always the transportation authority that selects the priority users. The selection has been based either on an arbitrary characteristic (like the license plate) that can vary from day to day (Daganzo and Garcia, 2000, Fosgerau, 2010, Knockaert *et al.,* 2016), or on the on-ramp at which users enter a highway (Lago and Daganzo, 2007, Shen and Zhang, 2010). The on-ramp based selection ramp facilitates the storage of parallel queues, but it is also possible to keep a pure temporal segregation with differentiated pricing (Daganzo and Garcia, 2000, Knockaert *et al.,* 2016).

This paper further develops the concept of temporal segregation by (i) laying a sound foundation for its study with heterogeneous users, (ii) analyzing different effects with heterogeneous users and (iii) leveraging the benefits of priority schemes to encourage carpooling. Section 2 introduces our modeling assumptions, further explains how temporal segregation could be implemented and provides some important preliminary results. Then, Sections 3, 4 and 5 all focus on different effects of temporal segregation. Section 3 revisits those known for homogeneous populations, Section 4 explores both the positive and negative consequences of heterogeneity in schedule preferences while Section 5 considers different types of vehicles to assess the additional benefits of prioritizing efficient modes (e.g. carpoolers). Finally, Section 6 combines all these effects and evaluates the distributional consequences of different types of dynamic priority schemes.

2 Modeling assumptions and some fundamental preliminary results

Vickrey's bottleneck problem considers departure time choice in the simplest possible setting: some total population of size $N > 0$ wants to go from A to B, and there is only one route connecting these two places. This section details the modeling assumptions used for both the demand and supply sides, and then addresses some fundamental issues like the existence and uniqueness of equilibria.

2.1 Demand model and assumptions

The demand model considered here is the rather general one proposed by Lindsey (2004):

Assumption 1. *The population consists of n homogeneous groups of users. Group i has a size N_i, a preferred arrival time t_i^* and a trip cost function $c_i(T, x) = \alpha_i T + D_i(x)$, where T denotes the extra travel time due to congestion, $\alpha_i > 0$ and $x = t - t_i^*$ represents the schedule delay. D_i is continuous, non-negative, satisfies $D_i(0) = 0$ and is such that for all feasible x,*

$$\lim_{\Delta x \to 0^+} (D_i(x + \Delta x) - D_i(x))/\Delta x > -\alpha_i.$$

The function D is known as the schedule penalty function. The widely-used $\alpha-\beta-\gamma$ preferences of Arnott *et al.* (1993) represent a simple example:

$$D(t - t^*) = \begin{cases} \beta(t^* - t), & \text{if } t < t^* \\ \gamma(t - t^*), & \text{otherwise.} \end{cases} \tag{1}$$

The coefficients β and γ account respectively for the costs of earliness and lateness. For convenience, we will also denote $N = \sum_{i=1}^{n} N_i$.

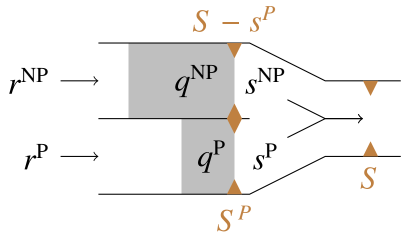


Figure 1: Schematic view with parallel queues

2.2 Supply model and assumptions

2.2.1 Mathematical description

The supply side in Vickrey's bottleneck problem is highly stylised: it considers a single bottleneck of constant capacity S, where all the queuing occurs in a FIFO manner, without explicit spatial propagation. The variant considered in this paper is similar to a highway merge with ramp metering. Users waiting upstream of the bottleneck are separated into two queues, denoted NP and P (see Fig. 1). Queue NP is open to all vehicles, while queue P is reserved for priority users. Similarly to a highway mainline, queue P is not metered. It has direct access to a subpart of the bottleneck with constant capacity $S^P \leq S$. If we denote $r^P(t)$ and $s^P(t)$ the flows of priority users upstream and downstream of the bottleneck and $q^P(t)$ the queue length, we have the classic bottleneck dynamics:

$$\left(\dot{q}^{\mathrm{P}}(t), s^{\mathrm{P}}(t)\right) = \begin{cases} \left(r^{\mathrm{P}}(t) - S^P, S^P\right), & \text{if } q^P(t) > 0 \text{ or } r^P(t) > S^P \\ \left(0, r^{\mathrm{P}}(t)\right), & \text{otherwise.} \end{cases} \tag{2}$$

Then, similarly to a metered ramp, the approach NP has a time-dependent capacity, which depends on the flow on the non-metered approach. If we define similarly $r^{\mathrm{NP}}(t)$ and $s^{\mathrm{NP}}(t)$ as the flows of non-priority users upstream and downstream of the bottleneck and $q^{\mathrm{NP}}(t)$ the number of non-priority vehicles queuing at time t, we have:

$$\left(\dot{q}^{\mathrm{NP}}(t), s^{\mathrm{NP}}(t)\right) = \begin{cases} \left(r^{\mathrm{NP}}(t) - (S - s^{\mathrm{P}}(t)), S - s^{\mathrm{P}}(t)\right) & \text{if } q^{\mathrm{NP}}(t) > 0 \text{ or } r^{\mathrm{NP}}(t) > (S - s^{\mathrm{P}}(t)) \\ \left(0, r^{\mathrm{NP}}(t)\right), & \text{otherwise.} \end{cases} \tag{3}$$

Thus, the metering scheme ensures that the bottleneck capacity is fully used whenever it is possible. There are usually very few time constraints in terms of passage time. Here, we simply define the two following alternative assumptions, with respectively constant and time-varying capacity.

Assumption 2. *The bottleneck is open with constant capacity $S > 0$ during a time interval*

$\bar{T} = (t_0, t_e)$*, such that* $\int_{t_0}^{t_e} S(t)\,\mathrm{d}t > N$ *and for all* $i = 1, ...n$*,* $t_i^* \in \bar{T}$*.*

Assumption 3. *The bottleneck is open with a piece-wise continuous capacity* $S(t) > 0$ *during an interval* $\bar{T} = (t_0, t_e)$*, such that* $\int_{t_0}^{t_e} S(t)\,\mathrm{d}t > N$ *and for all* $i = 1, ...n$*,* $t_i^* \in \bar{T}$*.*

2.2.2 Practical view

Note that the metered approach should have a physical capacity of at least S, while the capacity S^P of the priority approach can be smaller. Yet, these queues do not necessarily need to be at the same milepost on the highway, such that it is not actually necessary that the road upstream of the bottleneck has a capacity $S + S^P$. If the road with a bottleneck has m lanes and we accept to allocate to priority users at most all the lanes but one ($S^P = \frac{m-1}{m} S$)), Small (1983) proposed an ingenious scheme that stores queues of non-priority vehicles at different mileposts for each lane, with priority vehicles being allowed to slalom around these queues while non-priority vehicles would have to stay on their lane.

As another alternative to parallel queues, Daganzo and Garcia (2000) proposes to impose a fine toll on non-priority users that is just large enough to ensure that the two types of users never travel simultaneously (but without removing queuing altogether). Knockaert *et al.* (2016) follows a similar approach but considers instead a coarse toll. We do not consider such interpretations here as (i) they still involve some pricing, (ii) fine tolls are difficult to implement while coarse tolls theoretically lead to impractical departures masses and (iii) such strategies would ultimately lead to different equilibrium departure profiles.

2.3 Equilibrium: Definition, existence and uniqueness

In the present context, an equilibrium is a situation such that no individual user can reduce her travel cost by unilaterally changing departure time. Since the demand model is deterministic, this type of equilibrium is commonly referred to as Deterministic Dynamic User Equilibrium (DDUE). Equilibria are relevant insofar as they represent good approximations of the real world averages. Previous simulation results suggest that even though the DDUE might not be stable strictly speaking, the congestion cost approximation it provides are relatively good in realistic settings (heterogeneous users and reasonable rational adjustment mechanisms)(Lamotte, 2018).

Without priority, the existence and uniqueness of such an equilibrium were shown by Smith (1984) and Daganzo (1985) for populations having convex schedule penalty functions D and with a continuum of users differing only in t^*. Lindsey (2004) extended these results to account for non-convex and heterogeneous schedule penalty functions, as per Assumption 1. Iryo and Yoshii (2007) and Akamatsu *et al.* (2018) later provided another proof of existence relying on essentially the same assumptions as Lindsey (2004), but providing at the same time a method to construct the equilibrium. More specifically, Iryo and Yoshii (2007) and Akamatsu *et al.* (2018) showed the DDUE conditions are equivalent to solving a linear program, where the objective is to minimize a weighted sum of the schedule penalties of all groups, where the weight of group i is simply α_i^{-1}. The difference between these two papers is that Iryo and Yoshii (2007) assumes a discrete range of possible passage times, while Akamatsu *et al.* (2018) allows for a continuum.

Although the mathematical conditions for existence and uniqueness are not the main focus of the present paper, these results are required for the subsequent derivations. We thus propose the following adaptation of the results previously mentioned.

Proposition 1. *Let Assumptions 1 and 2 hold. Assume that a proportion $x \in (0, 1)$ of users of each group benefits from a metering-based priority scheme with a maximum capacity preemption $S^P \in (xS, S)$. In deterministic departure-time user equilibrium, the travel costs of each group are uniquely defined.*

The proof will be provided in the journal version of this paper.

3 Perfectly homogeneous users

This section assumes that all users are identical, but that a proportion $p \in [0, \bar{p}]$ benefits from metering-based priority. The maximum proportion that can be prioritized is defined by $\bar{p} = \frac{S^P}{S}$, such that priority users always have a favorable demand-to-capacity ratio.

3.1 Constructing the equilibrium

Equilibria under metering-based priority are most easily constructed sequentially. Indeed, the equilibrium departures of priority users can be determined as if they were on their own and non-priority users did not exist. This corresponds to a "normal" bottleneck problem with constant capacity S^P. Once the equilibrium exit flow of priority users s^P is known, it can be subtracted from the total capacity S, so that the equilibrium can be computed for the non-priority users.

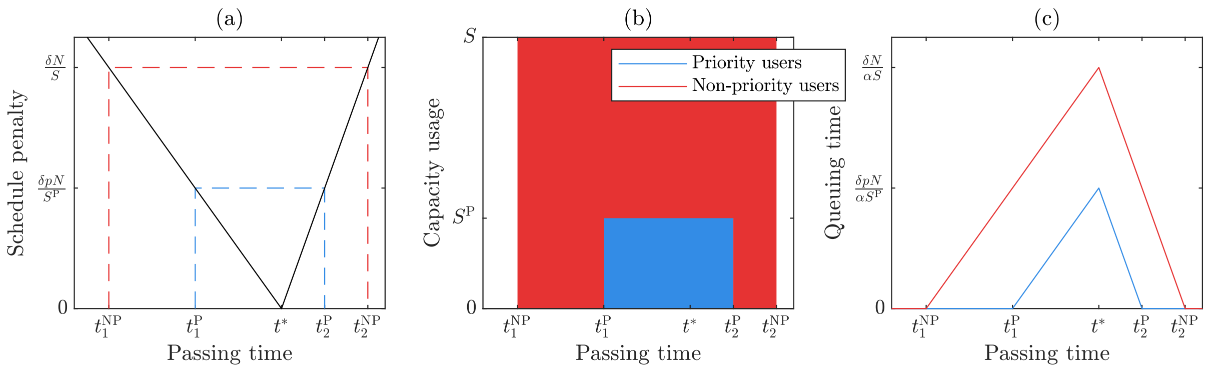


Figure 2: Equilibrium with metering-based priority and homogeneous users

We illustrate this method hereafter with the $\alpha - \beta - \gamma$ preferences and then come back to the general homogeneous case to provide expressions of the individual costs.

Fig. 2 represents a equilibrium with metering-based priority when users have $\alpha-\beta-\gamma$ preferences (with $\gamma = 2\beta$). Priority users only travel during the period $[t_1^P, t_2^P]$ which is defined by

$$\begin{cases} t_1^P < t^*, t_2^P > t^* \text{ and } D(t_1^P) = D(t_2^P) \\ (t_2^P - t_1^P)S^P = pN. \end{cases} \tag{4}$$

They pass the bottleneck at capacity during this interval, i.e. $s^P(t) = S^P\ \forall t \in [t_1^P, t_2^P]$. Together, the constraints in Eq. (4) imply that $D(t_1^P) = D(t_2^P) = \delta \frac{pN}{S^P}$, with $\delta = \frac{\beta\gamma}{\beta+\gamma}$, so the times t_1^P and t_2^P can be found as $(D|_{[t_0,t^*]})^{-1}\left(\delta \frac{pN}{S^P}\right)$ and $(D|_{[t^*,t_e]})^{-1}\left(\delta \frac{pN}{S^P}\right)$, where $D|_A$ denotes the restriction of D to A. The queuing time experienced by priority users passing the bottleneck at t is then such that it compensates the variations of $D(t)$ over the congested period:

$$T(t) = \begin{cases} \alpha^{-1}\left(\delta \frac{pN}{S^P} - D(t)\right), & \text{if } t \in [t_1^P, t_2^P] \\ 0, & \text{otherwise.} \end{cases}$$

If we now turn to non-priority users, their travelling period $[t_1^{NP}, t_2^{NP}]$ is defined by:

$$\begin{cases} t_1^{NP} < t^*, t_2^{NP} > t^* \text{ and } D(t_1^{NP}) = D(t_2^{NP}) \\ \int_{t_1^{NP}}^{t_2^{NP}} S - s^P(t)\,\mathrm{d}t = (1-p)N. \end{cases}$$

Since $\int_{t_1^{NP}}^{t_2^{NP}} s^P(t)\,\mathrm{d}t = pN$, the second equation reduces to $\int_{t_1^P}^{t_2^P} S\ \mathrm{d}t = N$. It implies that $D(t_1^{NP}) = D(t_2^{NP}) = \delta \frac{N}{S}$, i.e. t_1^{NP} and t_2^{NP} are the same as if nobody was prioritized. The queuing time is then

$$T(t) = \begin{cases} \alpha^{-1}\left(\delta\frac{N}{S} - D(t)\right), & \text{if } t \in [t_1^{NP}, t_2^{NP}] \\ 0, & \text{otherwise.} \end{cases}$$

Let us now come back to the general case, with an unspecified schedule penalty function. Let C^P and C^{NP} denote the equilibrium individual costs, and $C(N)$ be the function that maps a homogeneous population of size N with the associated individual equilibrium congestion cost when accessing a bottleneck of constant capacity S (without any priority scheme). This function, which depends on the schedule preferences, is known as the reduced form cost function. The relations that were found with the $\alpha - \beta - \gamma$ preferences between C^P, C^{NP} and C actually hold more generally, as shown hereafter.

Proposition 2. *Consider a homogeneous population satisfying Assumption 1 and a bottleneck satisfying 2. Under DDUE with metering-based priority,* $C^P = C\left(pN\frac{S}{S^P}\right)$ *and* $C^{NP} = C(N)$*. The DDUE with metering-based priority represents a Pareto Improvement compared to the DDUE with no priority.*

The proof is given Appendix A.1.

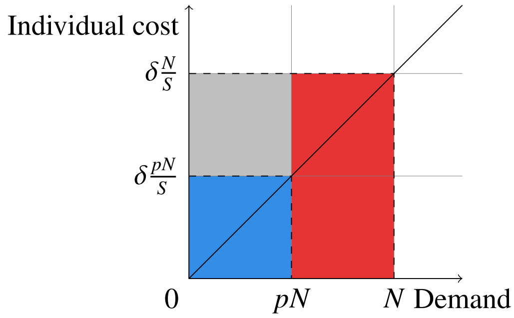
3.2 Optimal priority scheme

This section considers the effect of priority on the social cost,

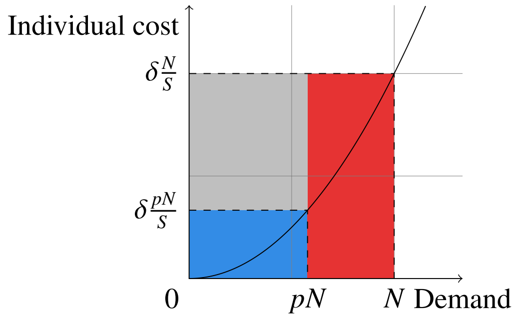
$$\mathrm{SC}(S^P, p) = pNC^P + (1-p)NC^{\mathrm{NP}}. \tag{5}$$

Since Proposition 2 implies that C^{P} reduces with S^P while C^{NP} remains constant, our simple model suggests that the social cost decreases with S^P, such that the limit case $S^P = S$ is socially optimal. In real life, the existence of unplanned trips or of some uncertainty in the bottleneck access time would require setting $S^P < S$. Such effects would however require more elaborate models and are considered beyond the scope of this paper.

The effect of the proportion of priority users p is not as simple. In the case of a linear cost function $C(N) = \delta\frac{N}{S}$, $\mathrm{SC}(S^P, p)$ is a second order polynomial, which is minimized for $p = \frac{S^P}{2S}$, such that the cost of priority users is exactly half the cost of the others. If we set $S^P = S$, we recover the situation studied by Daganzo and Garcia (2000) and illustrated in Fig. 3a. Graphically, the blue rectangle (of width pN and height $\delta\frac{pN}{S}$) represents the contribution of priority users to the social cost, while the red rectangle (of width $(1-p)N$ and height $\delta\frac{N}{S}$)



(a) Linear



(b) Quadratic

Figure 3: Linear and quadratic reduced cost functions, and their optimal priority schemes.

represents the one of non-priority users. Clearly, the sum of the areas of the two rectangles (i.e. the social cost) is minimized when $p = 0.5$ and it leads to a social cost reduction (the gray area) equal to 25% of the social cost without priority (which is the sum of the blue, red and gray areas).

The comparison with Fig. 3 suggests that metering-based priority schemes can provide even larger benefits when users have a convex cost function. This idea is formalized in the following proposition.

Proposition 3. *When applying metering-based priority to a homogeneous population having a continuously differentiable, strictly increasing and convex reduced form cost function, the optimal proportion of priority users is larger than $\bar{p}/2$, and it leads to an overall congestion cost reduction of at least $\bar{p}/4$.*

The conditions under which the reduced form cost function is convex are examined in Appendix A.2. It is easy to see that the reduced form cost function is convex when the schedule preferences satisfy Assumption 1 and are themselves convex, but this is not the only case.

Fosgerau (2011) showed a similar result for a different class of utility functions (strictly concave in arrival and departure time at bottleneck, increasing with arrival time and decreasing with departure time). Specifically, Fosgerau (2011) showed that for this class of utility functions, priority schemes achieve at least half the social cost reduction of the ideal fine toll. One can then recover the result in Proposition 3 by showing that with such schedule penalty functions, the ideal fine toll decreases the social cost by at least 50%. Since our utility function linearly decreases with the departure time, Fosgerau's specification admits ours as a limit case.

Let SC^{mbp} and SC^{ref} represent the social cost with and without priority. Fig. 4 shows the relative social cost reduction $(SC^{ref} - SC^{mbp})/SC^{ref}$ obtained with linear and quadratic cost functions for all the range of possible S^P/S and p. In the quadratic case, the maximum social

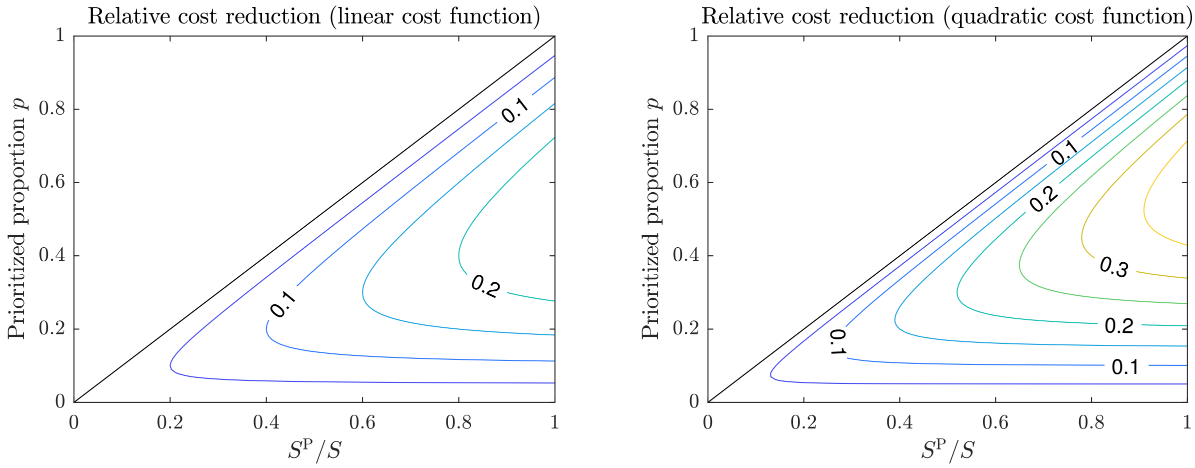


Figure 4: Relative reduction of the social cost obtained with metering-based priority for homogeneous users with linear (left) and quadratic (right) reduced form cost functions.

cost reduction is obtained for $(S^P, p) = (S, \sqrt{\frac{1}{3}})$, and it leads to $\mathrm{SC}(p^*) = aN^3\left(1 - \frac{2}{3\sqrt{3}}\right)$, i.e. a social cost reduction of about 39 %.

4 The contrasted effects of priority with heterogeneous users

This section studies the effects of metering-based priority on a population of users having different schedule preferences. Since schedule preferences are difficult to observe, we assume that the priority status is granted *independently* of them. Otherwise, this would open the way to a complex and probably inefficient game, in which users might act as if they had other schedule preferences, with the goal of being prioritized. This section shows with two simple examples how heterogeneity in schedule preferences might either reduce or increase the benefits of priority schemes.

4.1 The 2-flexibility example

The first example that we consider shows how priority may generate some inefficiencies by affecting the sequence in which users pass the bottleneck. In fact, the laissez-faire policy is known to minimize the sum of schedule penalties (Iryo and Yoshii, 2007) when users all have the same value of time α. Thus, if the priority status is allocated independently of user's preferences the resulting order is likely to yield greater schedule penalties than the one resulting from the laissez-faire equilibrium.

Table 1: Individual costs with flexible and inflexible users

	Without metering	With metering-based priority	
	All users	Priority users	Non-priority users
Flexible	$\delta_F(N_F + N_I)$	$\delta_F p(N_F + N_I)$	$\delta_F(N_F + N_I)$
Inflexible	$\delta_I N_I + \delta_F N_F$	$\delta_I pN_I + \delta_F pN_F$	$\delta_I(pN_F + N_I) + \delta_F(1-p)N_F$

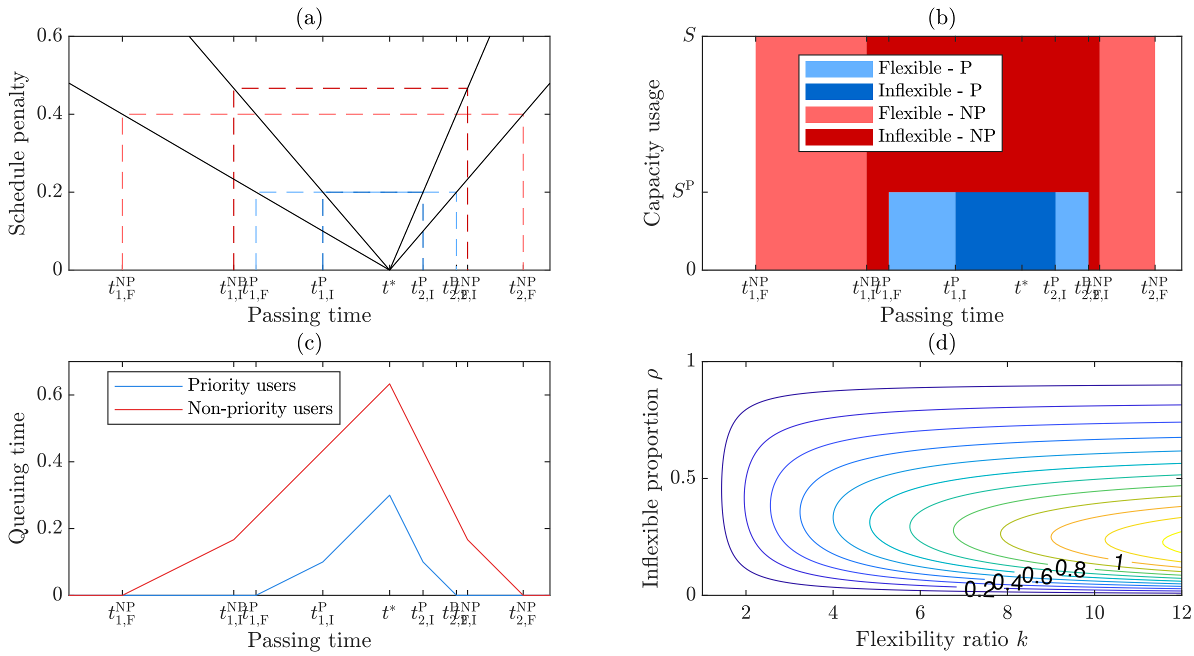


Figure 5: 2-flexibility case. (a-b-c): Representations of the schedule penalties (a), the capacity usage (b) and the queuing time (c) as functions of the passing time with $\rho = 0.5$, $k = 2$, $p = 1/6$ and $S^P = S/3$. (d): Contour plot of $H(k, \rho)$.

The example considered involves two homogeneous groups: one "flexible" and one "inflexible", of size N_F and N_I. Both groups have $\alpha - \beta - \gamma$ preferences with the same preferred arrival time t^*, the same value of time α, but the coefficients β and γ of inflexible users are k times ($k > 1$) as large as those of flexible users: $\beta_I = k\beta_F$ and $\gamma_I = k\gamma_F$. As in Section 2.1, let $\delta_F = \frac{\beta_F\gamma_F}{\beta_F+\gamma_F}$ and $\delta_I = \frac{\beta_I\gamma_I}{\beta_I+\gamma_I}$. Clearly, $\delta_I = k\delta_F$.

Without priority, the individual costs can be easily computed by following Arnott *et al.* (1993). With metering-based priority, we can again construct the equilibrium sequentially by applying the same approach first for priority users and then for the others. The various costs are summarized in Table 1 and the dynamics with metering are illustrated in Fig. 5abc.

Let us now analyze the impact of priority on social cost. To simplify, we focus in this analysis on the limit case $S^P = S$. If we denote $\mathrm{SC}^{\mathrm{mbp}}$ and $\mathrm{SC}^{\mathrm{ref}}$ the social costs with and without metering-based priority, the relative change in social cost is

$$\frac{\mathrm{SC^{mbp}} - \mathrm{SC^{ref}}}{\mathrm{SC^{ref}}} = \frac{p(N_F C_F^P + N_I C_I^P) + (1-p)(N_F C_F^{NP} + N_I C_I^{NP}) - N_F C_F^{ref} - N_I C_I^{ref}}{N_F C_F^{ref} + N_I C_I^{ref}}.$$

By replacing the individual costs by their values in Table 1, the relative change in social cost can be rewritten

$$\frac{\mathrm{SC^{mbp}} - \mathrm{SC^{ref}}}{\mathrm{SC^{ref}}} = p(1-p)(H-1), \tag{6}$$

where $H = \frac{(\delta_I - \delta_F) N_I N_F}{\mathrm{SC^{ref}}}$ captures the effect of heterogeneity in flexibility. Indeed, the population is homogeneous if and only if $H = 0$, in which case the problem reduces to the one studied in Section 3.2. Since A is non-negative, heterogeneity clearly reduces the benefits of priority in the 2-flexibility problem. If we further denote $\rho = \frac{N_I}{N_I + N_F}$, H can also be rewritten as

$$H(k,\rho) = \frac{(k-1)p_I(1-\rho)}{k\rho^2 + (1-\rho)^2 + 2\rho(1-\rho)}.$$

One can then show that for $k \leq 9$, $H(k,\rho) \leq 1$ for all $\rho \in [0,1]$, i.e. metering-based priority is always welfare improving, regardless of the exact value of p and ρ. If however $k > 9$, there are some ρ which lead to $H(k,\rho) > 1$, such that metering-based priority may decrease welfare. This is visible in Fig. 5d, which shows the variations of $H(k,\rho)$.

4.2 Uniformly distributed t^*

The second example shows how metering-based priority may potentially yield larger benefits than in the homogeneous case by ensuring that priority users experience no congestion at all. We consider a case where users have the same $\alpha - \beta - \gamma$ coefficients, but where t^* is uniformly distributed on an interval $[t_1^*, t_2^*]$ such that $(t_2^* - t_1^*) \triangleq \Delta^* \in \left[0, \frac{N}{S}\right)$. This set-up is not consistent with Assumption 1 because of its infinite number of group, but it is well covered by the literature (Vickrey, 1969) and it is probably the simplest and clearest way to illustrate the effect at hand.

The social cost decomposition at equilibrium is displayed in Fig. 6c as a function of Δ^*. The total queuing time is constant for $\Delta^* \in \left[0, \frac{N}{S}\right)$, but the total schedule penalty decreases linearly and converges to 0 as Δ^* tends towards $\frac{N}{S}$. The case $\Delta^* = \frac{N}{S}$ is degenerate[1], and the cases $\Delta^* > \frac{N}{S}$ exhibit no congestion at all, because the "demand density" (N/Δ^*) is everywhere smaller than the capacity S.

[1]There can be two equilibria, one with significant delays (the limit case as $\Delta^* \to \frac{N}{S}$ from below) and one with no delay at all.

The same two cases should be distinguished when considering the priority users of a metering-based priority scheme. Their demand density is pN/Δ^*, while their capacity is S^P. Priority users can thus experience no congestion at all if $p < \Delta^*/(N/S^P)$.

Case with no congestion for priority users ($p < \Delta^*/(N/S^P)$). Let us consider separately the sum of schedule penalties and the sum of queuing costs. The sum of schedule penalties over all users (both priority and non-priority) remains equal in all the scenarios considered. Indeed, the arrival orders of these scenarios can be obtained by reallocating the passage times of early (resp. late) or on-time users in such a way that all these users remain early (resp. late) or on time, and it is easy to see that with homogeneous β and γ coefficients, such modifications leave the sum of schedule penalties unchanged.

Let us now consider the graph of queuing costs in Fig. 6a. The queuing time profile for non-priority users is the same as without metering, and overall, the distribution of passage times at the bottleneck remains the same (uniformly distributed between t_1 and t_2). The only difference is that pN users uniformly distributed between t_1^* and t_2^* experience no delay at all. It is easy to see that the average queuing time corresponding to these passage times would have been the average of $\delta\frac{N}{S}$ (the maximum) and $\delta\left(\frac{N}{S} - \Delta^*\right)$ (the minimum). Thus, the total queuing cost saved is $pN\left(\frac{N}{S} - \frac{\Delta^*}{2}\right)$. Since the overall congestion cost without metering is $\delta N\left(\frac{N}{S} - \frac{\Delta^*}{2}\right)$, metering-based priority simply reduces the overall congestion cost by a proportion p.

Case with congestion for all users ($p > \Delta^*/(N/S^P)$). As in the previous case, the sum of schedule penalties remains unchanged. Here however, priority users do experience queuing at equilibrium, as in Section 3. An example of such scenario is shown in Fig. 6c. The pN priority users traveling between t_1^P and t_2^P benefit from a reduction of $\delta\left(\frac{N}{S} - \frac{pN}{S^P}\right)$ in their individual queuing cost. Thus, the overall social cost reduction is

$$\frac{SC^{ref} - SC^{mbp}}{SC^{ref}} = \frac{p\left(\frac{N}{S} - \frac{pN}{S^P}\right)}{\left(\frac{N}{S} - \frac{\Delta^*}{2}\right)}.$$

The relative social cost reductions achieved with and without congestion for priority users are shown together in Fig. 6d, as functions of p and $\Delta^*/(N/S)$ for the case $S^P = S$. The discontinuity for $p = \Delta^*/(N/S)$ and the upper bound of 100 % for the relative social cost reduction result from the uniform distribution assumption, which is admittedly not so realistic. Yet, this simple example illustrates two important results. First, priority can achieve very large benefits by allowing some users to travel without congestion. These benefits may largely outweigh those

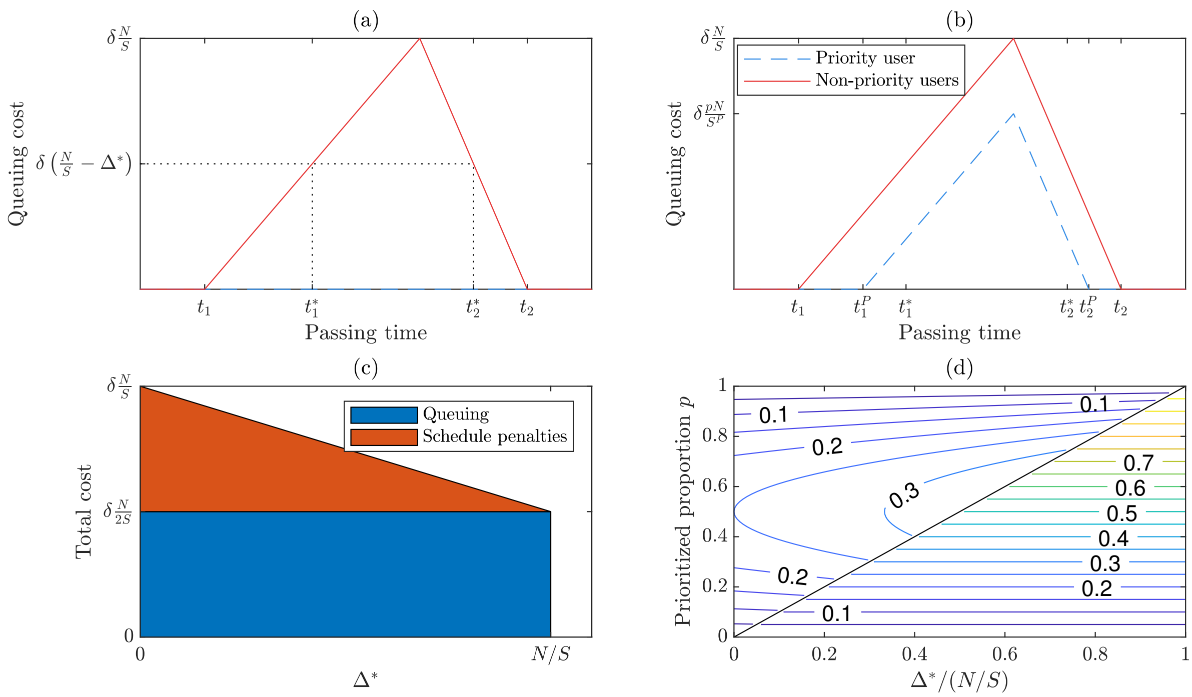


Figure 6: Uniformly distributed t^*. (a,b): Examples of queuing cost profiles with (a) $p < \Delta^*/(N/S^P)$ and (b) $p > \Delta^*/(N/S^P)$. (c): Total cost decomposition depending on the range of desired arrival times. (d): Relative social cost reduction obtained by metering with $S^P = S$.

found in the homogeneous case. Second, even when priority users are too numerous and have to queue, the relative social cost reduction may exceed the value found with homogeneous users (25 % in this case). This is because priority affects the queuing costs, and that these typically represent a larger proportion of the social cost when users have different preferences (see Fig. 6c).

5 The additional benefits of prioritizing efficient modes

This section comes back to the case of a population with homogeneous schedule preferences, but introduces two types of users: some "normal" users with a reference capacity usage of 1 and some "socially efficient" users with a capacity usage of $1/g$, with $g > 1$. An important difference with Section 4 is that the capacity usage is considered to be observable, so that it can be used as a criterion for the allocation of the priority status.

There could be plenty of interpretations for the type of vehicle used by "socially efficient" users: short cars, autonomous cars, motorbikes, carpools, etc. Yet, for the sake of clarity, we will think of efficient users as carpoolers. The coefficient g then represents the average occupancy of carpools, and prioritizing only efficient users then boils down to creating a metering-based HOV lane.

Note that mathematically, since we have made the continuum approximation, considering users with different capacity usages is equivalent to considering that some users (the efficient ones) have schedule preferences and values of time g times larger than others. The practical consequences are however very different as (i) prioritizing the richest users is economically regressive and (ii) it does not allow fostering virtuous mode choices. The remaining of this section explores the consequences of prioritizing socially efficient users, first with an exogenous proportion of efficient users (denoted q) and then when the proportion of efficient users depends on the benefits that this choice yields.

5.1 Exogenous mix

If we keep the same population size, the peak hour now lasts for $\left(1 - q + \frac{q}{g}\right)\frac{N}{S}$, such that the reference cost (without metering) is $C^{\text{ref}} = C\left(\left(1 - q + \frac{q}{g}\right)\frac{N}{S}\right)$. Obviously, the duration and the reference cost decrease with both g and q. More interesting is the influence of the two following priority schemes:

- Efficiency Priority (EP): all efficient vehicles are prioritized. The condition on q for priority vehicles to be indeed prioritized becomes $\frac{q}{gS^P} < (1 - q + \frac{q}{g})S^{-1}$. Provided that it is satisfied, the average cost is

$$C^{EP} = qC\left(\frac{qN}{gS^P}\right) + (1 - q)C^{\text{ref}}.$$

- Random Priority (RP): a proportion $p = q$ is prioritized, but the priority users are chosen randomly, independently of their efficiency. The average cost is then:

$$C^{RP} = pC\left(\frac{pN}{S^P}\left(1 - q + \frac{q}{g}\right)\right) + (1 - p)C^{\text{ref}}.$$

Note that the RP scheme defined above is equivalent to metering-based priority with a homogeneous population: if we consider two situations corresponding to the same congested period but with either homogeneous users or users having different efficiency, the relative social cost reduction obtained with RP metering is the same.

One can also already note that $1 - q + \frac{q}{g} = \frac{g(1-q)+q}{g} > \frac{1-q+q}{g} = \frac{1}{g}$, so that $C^{EP} \leq C^{RP} \leq C^{\text{ref}}$, i.e. the EP scheme always provides a greater relative social cost reduction than the RP scheme. The magnitude of this difference then depends on the prioritized proportion (q in the EP scheme, p

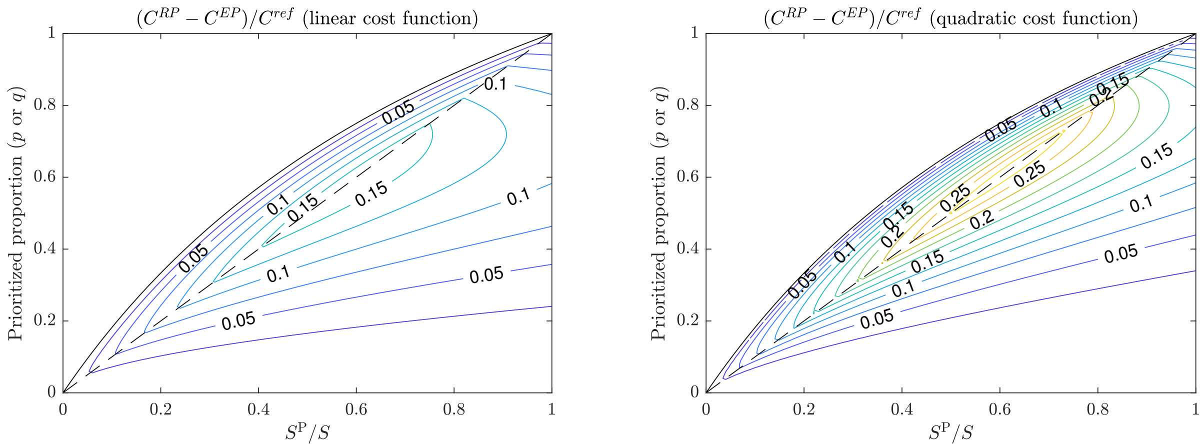


Figure 7: Additional relative social cost reduction obtained by prioritizing efficient users (EP scheme) instead of random users (RP), with $g = 2$.

in the RP one), the efficiency g and the capacity ratio S^P/S. Fig. 7 provides some illustration for the case $g = 2$ with both linear and quadratic cost functions.

The black solid and dashed lines indicate the feasible space frontiers, so that priority users are actually prioritized. This space is defined by $p < \bar{p} = S^P/S$ with the RP scheme and $q < \bar{q} = \frac{gS^P}{S+(g-1)S^P}$ with the EP scheme. Since $S^P/S < 1$ implies $\frac{g}{1+(g-1)\frac{S^P}{S}} > \frac{g}{1+(g-1)} = 1$, the EP scheme allows prioritizing a larger proportion of users than the RP scheme.

Fig. 7 suggests that the additional benefit of the EP scheme compared to the RP one can be very significant. For a given proportion of efficient users q, a given prioritized proportion and a given ratio S^P/S, the additional benefits may be of the same order of magnitude as the original benefits of the RP scheme, in particular when the number of prioritized users is large. This last observation can be supported by a simple analysis of the ratio of social cost reductions in the limit case $S^P = S$. Let us assume that the cost function is continuously differentiable, and let us consider the ratio $(C^{\mathrm{ref}} - C^{RP})/(C^{\mathrm{ref}} - C^{EP})$ as the prioritized proportion tends towards 0 and 1. To be fair, we should only compare situations with the same proportion of efficient users q, so we assume that $p = q$ in the RP scheme. The ratio of social cost reductions then reduces to

$$\frac{C^{\mathrm{ref}} - qC\left(\frac{qN}{S}\left(1-q+\frac{q}{g}\right)\right) - (1-q)C^{\mathrm{ref}}}{C^{\mathrm{ref}} - qC\left(\frac{qN}{gS}\right) - (1-q)C^{\mathrm{ref}}} = \frac{C^{\mathrm{ref}} - C\left(\frac{qN}{S}\left(1-q+\frac{q}{g}\right)\right)}{C^{\mathrm{ref}} - C\left(\frac{qN}{gS}\right)}.$$

On one hand, both $C^{\mathrm{ref}} - C(q(1-q+\frac{q}{g}))$ and $C^{\mathrm{ref}} - C(\frac{q}{g})$ tend towards C^{ref} as q tends towards 0, so their ratio tends towards 1. On the other, for q close to 1, the numerator is equal to $C'(\frac{1}{g})\frac{1-q}{g} + o(1-q)$, while the denominator is equal to $C'(\frac{1}{g})(1-q) + o(1-q)$.[2] Their ratio tends towards $\frac{1}{g}$. It suggests that if only a small proportion is prioritized, it is not crucial to base the

[2]We use here the "small o" notation: $u(x) = o(v(x))$ close to y if $u(x)/v(x) \to 0$ when $x \to y$.

selection on the traffic efficiency. Any other politically acceptable criterion would also be fine, as long as the mix of vehicles can be considered independent of the priority scheme. If however a large proportion is to be prioritized, prioritizing efficient users would yield significantly larger benefits.

5.2 Endogenous mix

Let us now focus on the case where only "efficient" users are prioritized, and where the proportion of such users is endogenous. With the interpretation of efficient users as carpoolers, users would choose to carpool if and only if the advantage they derive from being prioritized outweighs the inconvenience cost of carpooling.

5.2.1 A distribution of inconvenience cost

We assume that all other things being equal, users have a preference for either the efficient mode, or for the normal one. If we take the normal mode as a reference, all users have a personal inconvenience cost $\chi \in \mathbb{R}$ associated to being efficient, which can be positive or negative. In the case of carpooling, χ would account for a variety of inconveniences such as the need to detour, the privacy loss, or the extra organizational load, as well as for carpooling advantages, such as the opportunity to socialize or the sharing of fixed costs related to traveling.

Let us assume that χ follows a continuous distribution and let f and F denote its pdf and cdf. The following assumption guarantees that even without any priority scheme, a proportion $F(0) > 0$ of users carpool.

Assumption 4. *The support of the probability density function of the carpooling cost is an interval including 0 and not bounded above.*

5.2.2 Equilibrium mix

Since users are assumed to differ only in their inconvenience cost χ, the equilibrium efficient proportion q is simply $q = F(a)$, where a denotes the advantage users derive from being prioritized. Yet, for a given preemptable capacity S^P, the queuing advantage experienced by priority users depends on the number of priority users. Specifically, if there are $q \in [0, 1]$

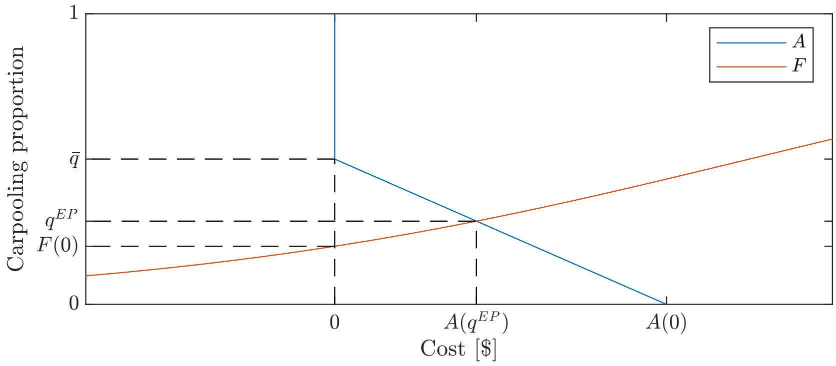


Figure 8: Equilibrium with endogenous proportion of efficient users, with $\alpha-\beta-\gamma$ preferences, a logistic distribution of the carpooling cost χ and with $S^P = S/3$.

efficient users, their congestion advantage is

$$A(q) = \begin{cases} C\left(\left(1 - q + \frac{q}{g}\right)\frac{N}{S}\right) - C\left(\frac{qN}{gS^P}\right), & \text{if } q \in [0, \bar{q}]\,, \\ 0, & \text{if } q \in [\bar{q}, 1]\,. \end{cases}$$

Thus, solving for the equilibrium reduces to solving the fixed point problem $q = F(A(q))$.

Proposition 4. *Let assumptions 1 and 2 hold. Assume that the population differs only in the carpooling inconvenience cost, that Assumption 4 holds, and that carpoolers benefit from a metering-based priority scheme with $S^P < S$. There exists a unique equilibrium proportion of carpoolers q^{EP}. It is strictly greater than $F(0)$ is and only if $F(0) < \bar{q}$.*

This result is relatively obvious graphically (Fig. 8), but its proof is provided in Appendix. The magnitude of the increase in the carpooling proportion strongly depends on the distribution of the carpooling cost and on congestion severity. When congestion is very severe, such a priority treatment is equivalent to providing a very significant monetary reward to carpoolers. Note also that even when such a scheme is implemented, the proportion of carpoolers remains smaller than it should be to minimize the social cost. Indeed, users do not account for the overall positive externality they would have on others if they carpooled.

6 Case-study

6.1 Description

This section presents numerical results regarding the social cost and distributional consequences of the various schemes discussed in the paper. Since the number of parameters is too large

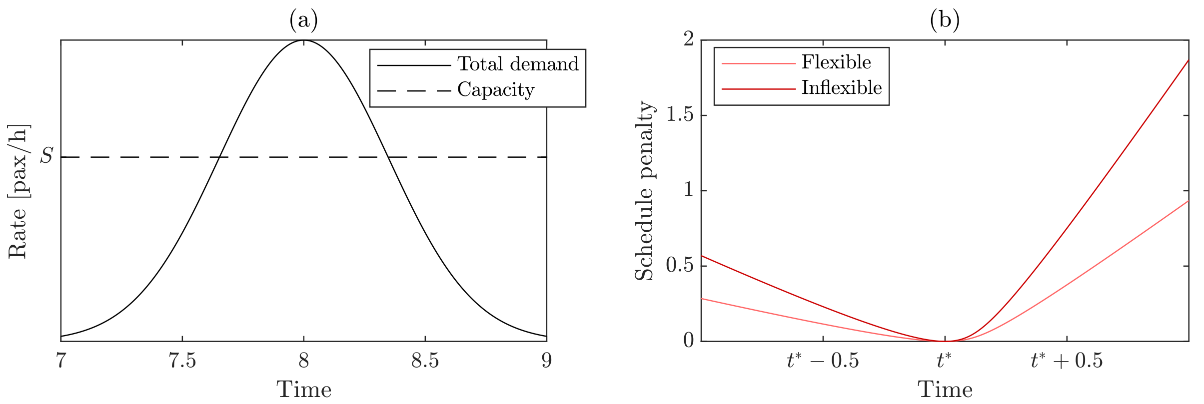


Figure 9: Demand description: (a) distribution of desired passage times and (b) schedule penalty functions.

for a complete sensitivity analysis, we focus on a single scenario designed to be realistic and combining all sources of heterogeneity.

As in Section 4.1, half of the users have a schedule penalty function k times larger than the other half (with $k = 2$). These are illustrated in Fig. 9b. The functional form used is an approximation of the $\alpha-\beta-\gamma$ preferences, given by $D(t-t^*) = \int_{t^*}^{t} \frac{\delta}{\pi} \tan^{-1}(w(s-\tilde{t}))\,\mathrm{d}s$, with $\tilde{t} = t^* + \tan(\frac{\pi(\gamma-\beta)}{2(\gamma+\beta)})\frac{1}{w}$ and $w = 10$ h^{-1}. Users also have heterogeneous desired arrival times t^*, but these are now normally distributed for more realism (with a mean $t = 8$ h and a standard deviation of 0.35 h). The ratio N/S was then set to 0.7, such that the total desired passage rate exceeds capacity between approximately 7h40 and 8h20, as shown in Fig. 9a.

The carpooling cost χ is assumed to follow a logistic distribution. We parameterized it such that $F(0) = 20\%$ (in line with trends reported in the literature), and then tried different values for the mean carpooling cost $\bar{\chi}$, ranging from 0.4 h to 10 h. Based on a short literature review, a reasonable value would seem to be around 0.5 h. Finally, the maximum preemptable capacity S^P was taken equal to $S/3$.

6.2 Computational method

The method used to compute equilibria is based on Iryo and Yoshii (2007), and it consists in solving an equivalent Linear Program (LP). This method assumes a finite set of groups (our n) and a finite set of possible bottleneck passage times. Here, we considered a set of times between 7 h and 9 h with a regular spacing of $1/200$ h (400 possible passage times in total). The distribution of desired arrival time was discretized on the same set, so that the reference scenario included 2 sets (one flexible and one inflexible) of 400 groups each, with normally distributed group sizes N_i. The cases with metering-based priority were solved by applying the

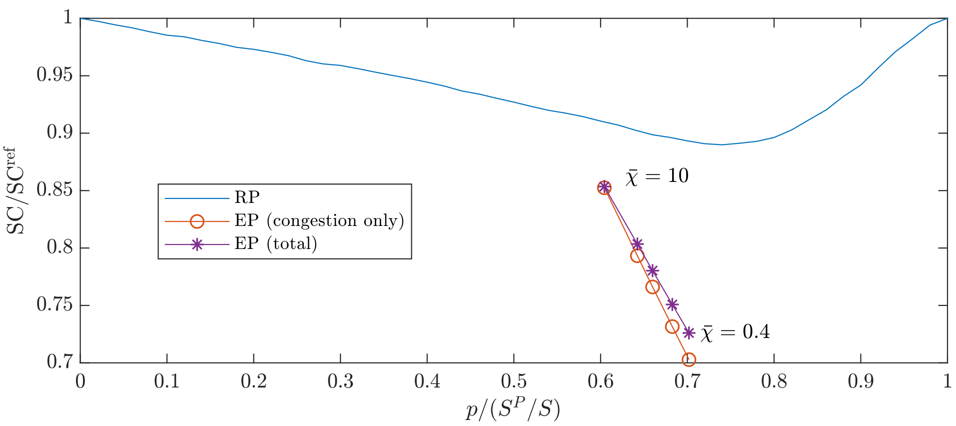


Figure 10: Relative social cost with Random Priority (p varies between 0 and S^P) and with Efficiency Priority (with average carpooling cost $\bar{\chi}$ equal to 0.4, 0.5, 0.7, 1 and 10).

same algorithm twice (first for the priority users, then for the others).

6.3 Social cost comparison

We are interested in how, starting from the reference scenario described in Section 6.1, metering-based priority can reduce the cost of congestion. This section takes a global perspective and looks only at the social cost (i.e. the sum of all individual costs), while the next one looks at the distributional consequences (at the individual level).

Fig. 10 represents the social cost in a broad range of situations. The blue curve shows how it varies with the prioritized proportion, when this proportion is chosen randomly (see Section 5.1). Here, the largest social cost reduction (11%) is obtained for $p = 0.74\frac{S^P}{S}$. The magnitude of this reduction is consistent with the 2-flexibility example of Section 4.1, while the prioritized proportion is larger than $O.5\frac{S^P}{S}$, in agreement with Proposition 3 (for the homogeneous case). Then, the blue and purple curves show the social cost reduction obtained when prioritizing carpoolers only, with an endogenous mix. The red curve shows the total congestion cost (queuing and schedule penalty), while the purple curve also includes the carpooling cost of the users that do not naturally carpool. The point with $\bar{\chi} = 10$ approximates the case with an exogenous proportion of carpoolers of $F(0)$. The average carpooling cost is so large in this case that the benefits of the metering-based priority scheme are not sufficient to create a significant mode shift. The comparison between the point $\bar{\chi}$ and its vertical projection on the blue line illustrates the importance of prioritizing efficient users, even with a given mix (as explained in Section 5.1).

The other points on the blue and red lines then illustrate the additional benefits gained by incentivizing carpooling. These can greatly increase the social cost reduction, even with

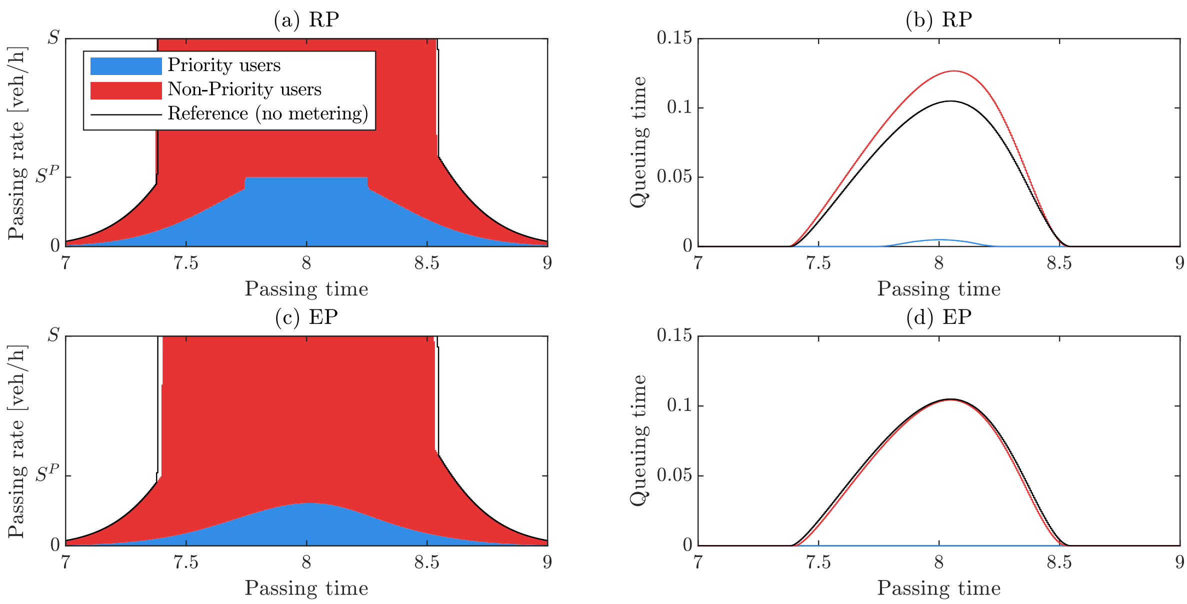


Figure 11: Passing rate and delays for the best RP scheme and for EP with $\bar{\chi} = 0.5$

relatively small mode shifts. The case with $\bar{\chi} = 0.4$ achieves a social cost reduction of 27.4 % (12.75 % more than the $\bar{\chi} = 0.4$) with 23.4 % of carpoolers (i.e. only 3.4 % more than in the reference scenario and in the case $\bar{\chi} = 10$).

6.4 Distributional consequences

We now compare more carefully three situations:

- the reference scenario without metering,
- the best RP scheme ($p = 0.74\frac{S^P}{S}$),
- a realistic EP scheme ($\bar{\chi} = 0.5$).

The passing rates and the queuing times for priority and non-priority users are shown in Fig. 11. As the RP scheme forces non-priority users to arrive further from their desired arrival time, the convexity of the schedule penalty function leads to steeper travel time variations at equilibrium. In the end, this results in a queuing profile for non-priority users that exceeds the reference one. A similar effect also exists with the EP scheme, but it is counterbalanced by the mode shift towards carpooling, which slightly reduces the congested period duration. As a result, the queuing time experienced by non-priority users is always smaller than in the reference case.

This has significant consequences at the individual level. Fig. 12a shows the distribution of individual cost reductions when comparing the EP and RP scheme to the reference scenario. With the RP scheme, slightly more than 20 % of users are better-off, about 15 % are indifferent,

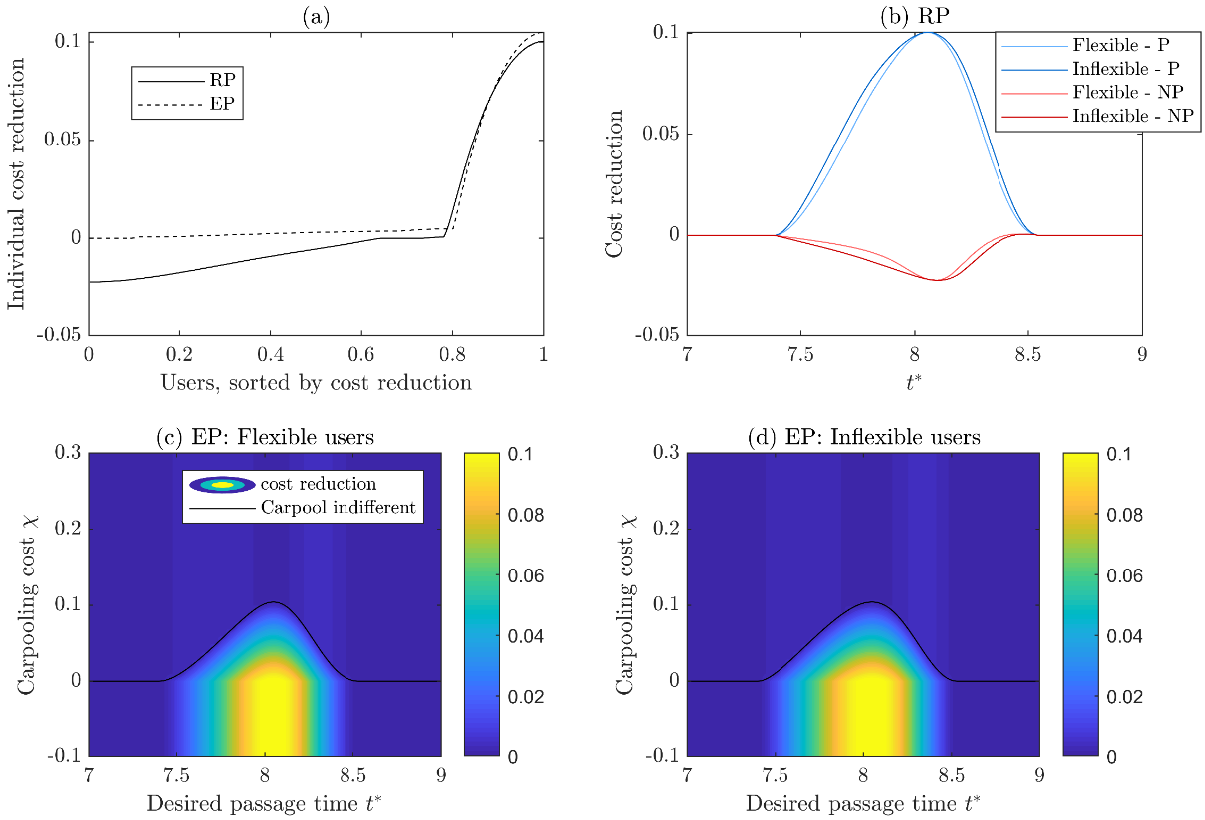


Figure 12: Distributional consequences of the best RP scheme and of the EP scheme with $\bar{\chi} = 0.5$: (a) global view of the individual cost reductions sorted in increasing order for both schemes, (b,c,d) individual cost reduction, per user type.

and the rest is worse-off. Fig. 12b provides information regarding the individuals that win and lose. Unsurprisingly, the worse off users are the non-prioritized ones who would like to pass the bottleneck in the middle of the peak. Thus, the social cost reduction of 11 % hides very different situations at the individual level. The existence of a large majority that is worse off may seriously compromise the feasibility of such a scheme.

On the other hand, Fig. 12a shows that with the EP scheme, all users are either better off or indifferent. The users that gain the most are naturally those that carpool (see Fig. 12c and d), but some others also have small gains, thanks to the induced mode shift.

7 Conclusive remarks

Metering-based priority is a very promising congestion alleviating measure, especially when its advantages are used to leverage carpooling or other efficient modes. As our numerical case-study showed, it has the potential to induce Pareto-improvements, even with very heterogeneous users. The inherently adaptive nature of its metering-based mechanism also confers to it a strong robustness. In fact, implementing a metering-based mechanism does not even require precise estimates of travelers' inclination towards carpooling. While a real-world implementation would be most valuable, another worthy research avenue would be to investigate whether similar priority schemes can be developed for other congestion mechanisms.

8 References

Akamatsu, T., K. Wada, T. Iryo and S. Hayashi (2018) Departure time choice equilibrium and optimal transport problems, *Working paper*. Accessed 2019-04-16.

Arnott, R., A. de Palma and R. Lindsey (1993) A structural model of peak-period congestion: A traffic bottleneck with elastic demand, *The American Economic Review*, **83** (1) 161–179.

Daganzo, C. F. (1985) The uniqueness of a time-dependent equilibrium distribution of arrivals at a single bottleneck, *Transportation science*, **19** (1) 29–37.

Daganzo, C. F. and R. C. Garcia (2000) A Pareto improving strategy for the time-dependent morning commute problem, *Transportation Science*, **34** (3) 303–311.

Dahlgren, J. (1998) High occupancy vehicle lanes: Not always more effective than general purpose lanes, *Transportation Research Part A: Policy and Practice*, **32** (2) 99–114.

Fosgerau, M. (2010) On the relation between the mean and variance of delay in dynamic queues with random capacity and demand, *Journal of Economic Dynamics and Control*, **34** (4) 598–603.

Fosgerau, M. (2011) How a fast lane may replace a congestion toll, *Transportation Research Part B: Methodological*, **45** (6) 845–851.

Iryo, T. and T. Yoshii (2007) Equivalent optimization problem for finding equilibrium in the bottleneck model with departure time choices, paper presented at the *4th IMA International Conference on Mathematics in Transport*, 231–244.

Knockaert, J., E. T. Verhoef and J. Rouwendal (2016) Bottleneck congestion: Differentiating the coarse charge, *Transportation Research Part B: Methodological*, **83**, 59–73.

Konishi, H. and S. Mun (2010) Carpooling and congestion pricing: Hov and hot lanes, *Regional Science and Urban Economics*, **40** (4) 173 – 186.

Lago, A. and C. F. Daganzo (2007) Spillovers, merging traffic and the morning commute, *Transportation Research Part B: Methodological*, **41** (6) 670–683.

Lamotte, R. (2018) Congestion and departure time choice equilibrium in urban road networks, Ph.D. Thesis, EPFL.

Lindsey, R. (2004) Existence, uniqueness, and trip cost function properties of user equilibrium in the bottleneck model with multiple user classes, *Transportation Science*, **38** (3) 293–314.

Shen, W. and H. Zhang (2010) Pareto-improving ramp metering strategies for reducing congestion in the morning commute, *Transportation Research Part A: Policy and Practice*, **44** (9) 676–696.

Small, K. A. (1983) Bus priority and congestion pricing on urban expressways, *Research in Transportation Economics*, **1**.

Small, K. A., C. Winston and J. Yan (2006) Differentiated road pricing, express lanes, and carpools: Exploiting heterogeneous preferences in policy design, *Brookings-Wharton Papers on Urban Affairs*, 53–96.

Smith, M. J. (1984) The existence of a time-dependent equilibrium distribution of arrivals at a single bottleneck, *Transportation Science*, **18** (4) 385–394.

Vickrey, W. S. (1969) Congestion Theory and Transport Investment, *American Economic Review*, **59** (2) 251–60.

A Individual costs and reduced form cost function

A.1 Proof of Proposition 2

The following lemma provides a general expression for the reduced form cost function.

Lemma 1. *Consider a problem satisfying Assumption 1 and 3, but with homogeneous population ($n = 1$). In deterministic departure-time user equilibrium, the individual cost is*

$$C(N) = \sup\{c \in SP \mid \int_{t_0}^{t_e} [D(t - t^*) < c] S(t)\, \mathrm{d}t \leq N\}, \tag{7}$$

where SP denotes the image of the set (t_0, t_e) under $t \mapsto D(t^* - t)$.

Proof. Let $s(t)$ denote the flow passing the bottleneck at time t. Equilibrium requires that no user can become better-off by changing departure time. Thus, all times t such that $D(t - t^*) < C(N)$ must satisfy $T(t) > 0$, and therefore $s(t) = S(t)$. By integrating over time and using the Iverson bracket notation[3], we have that $C(N)$ must be part of the set

$$Z_1 = \{c \in \mathbb{R} \mid \int [D(t - t^*) < C(N)] S(t)\, \mathrm{d}t \leq N\}. \tag{8}$$

Similarly, since the queuing time T cannot be strictly negative, all times such that $s(t) > 0$ must satisfy $D(t - t^*) \leq C(N)$. By integrating over time, $C(N)$ must also be part of the set

$$Z_2 = \{c \in \mathbb{R} \mid \int [D(t - t^*) \leq C(N)] S(t)\, \mathrm{d}t \geq N\}. \tag{9}$$

Besides, by assumption 3, we can take $t \in (t_0, t_e)$ such that $s(t) < S(t)$, and therefore $T(t) = 0$ and $D(t - t^*) \geq C(N)$. Since $N > 0$, we can find t' such that $s(t') > 0$ and therefore $D(t - t^*) \leq C(N)$. Since $t \mapsto D$ is continuous over the closed interval defined by t and t', we obtain that $C(N) \in$ SP.

Let us now assume that we have $c_1 \in Z_1 \cap$ SP, $c_2 \in$ SP and $c_2 < c_1$. Clearly,

$$\int_{t_0}^{t_e} [D(t - t^*) \leq c_2] S(t)\, \mathrm{d}t = \int_{t_0}^{t_e} [D(t - t^*) < c_1] S(t)\, \mathrm{d}t - \int_{t_0}^{t_e} [D(t - t^*) \in (c_2, c_1)] S(t)\, \mathrm{d}t.$$

Since $c_1 \in Z_1$, $\int_{t_0}^{t_e} [D(t - t^*) < c_1] S(t)\, \mathrm{d}t \leq N$, and since both c_1 and c_2 belong to SP and D is

[3][P] is equal to 1 if P is true and 0 otherwise.

continuous, $\int_{t_0}^{t_e}[D(t-t^*) \in (c_2, c_1)]S(t)\,\mathrm{d}t > 0$. Thus, $\int_{t_0}^{t_e}[D(t-t^*) \leq c_2]S(t)\,\mathrm{d}t < N$, i.e. $c_2 \notin Z_2$. The contrapositive is that all elements of $Z_2 \cap \mathrm{SP}$ are greater than all elements of $Z_1 \cap \mathrm{SP}$. Since $C(N)$ belongs to both, it is necessarily the supremum of $Z_1 \cap \mathrm{SP}$. □

Since priority users do not compete with non-priority ones, the individual cost they incur is simply $c^P = C(p\frac{N}{S^P})$. Non-priority users then compete among themselves for the remaining time-dependent capacity, $S - s^P(t)$. Using Lemma 1, their individual cost is

$$\begin{aligned} c^{\mathrm{NP}} &= \sup\{c \in \mathrm{SP} \mid \int_{t_0}^{t_e}[D(t-t^*) < c](S - s^P(t))x\,\mathrm{d}t \leq (1-p)N\} \\ &= \sup\{c \in \mathrm{SP} \mid \int_{t_0}^{t_e}[D(t-t^*) < c]S\ \mathrm{d}t - \int_{t_0}^{t_e}[D(t-t^*) < c]s^P(t)\,\mathrm{d}t \leq (1-p)N\}. \end{aligned}$$

Since $s^{\mathrm{P}}(t) > 0 \Rightarrow D(t-t^*) \leq c^P$ and $c^P < c^{\mathrm{NP}}$, $\int_{t_0}^{t_e}[D(t-t^*) < c]s^P(t)\,\mathrm{d}t = \int_{t_0}^{t_e} s^P(t)\,\mathrm{d}t = pN$. Thus,

$$c^{\mathrm{NP}} = \sup\{c \in \mathrm{SP} \mid \int_{t_0}^{t_e}[D(t-t^*) < c]S\ \mathrm{d}t \leq N\} = C(N).$$

A.2 Conditions for a convex reduced form cost function

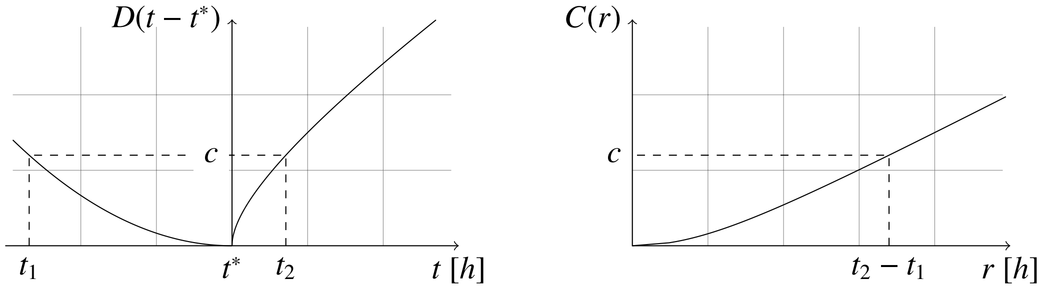
The general form of $C(N)$ provided by Eq. (7) implies that C is non-decreasing, but it is not very intuitive. If however we further assume that D is strictly decreasing for negative x and strictly increasing for positive x, Eq. (7) reduces to

$$\int [D(t-t^*) \leq C(N)]S(t)\,\mathrm{d}t = N.$$

With the $\alpha-\beta-\gamma$ preferences introduced in Section 2.1 and a constant capacity S, the solution to this equation is simply $C(N) = \delta\frac{N}{S}$, where $\delta = \frac{\beta\gamma}{\beta+\gamma}$.

Let us now examine the conditions under which C is convex (or concave). To obtain some intuition, we focus on the case of a bottleneck of constant capacity S and on a situation like the one illustrated in Fig. 13(a), where $t \mapsto D(t-t^*)$ is continuously differentiable on $(t_0, t^*) \cup (t^*, t_e)$, with $D'(t-t^*) < 0$ for early arrivals ($t < t^*$) and $D'(t-t^*) > 0$ for late arrivals ($t > t^*$).

In such a situation, the congested period at equilibrium always consists of a single interval (t_1, t_2). If the demand is sufficiently small so that t_1 and t_2 are different from t_0 and t_e, these bounds can be defined as functions of the equilibrium cost c as the only times in (t_0, t^*) and (t^*, t_e) satisfying $D(t_1(c) - t^*) = D(t_2(c) - t^*) = c$. Differentiating this with respect to c leads



(a) Schedule penalty function (b) Reduced form cost function

Figure 13: A non-convex schedule penalty and its a convex reduced cost function.

to $t_1'(c) = 1/D'(t_1 - t^*)$ and $t_2'(c) = 1/D'(t_2 - t^*)$. Note also that $C^{-1}(c) = N = S(t_2(c) - t_1(c))$. Combining these results leads to:

$$\begin{aligned}
C'(N) &= (C^{-1\prime}(C(N)))^{-1} \\
&= S^{-1}(t_2'(C(N)) - t_1'(C(N)))^{-1} \\
&= S^{-1}\left((D'(t_2(C(r))))^{-1} - (D'(t_1(C(r))))^{-1}\right)^{-1}.
\end{aligned}$$

If $t \mapsto D(t - t^*)$ is convex (resp. concave) on both (t_0, t^*) and (t^*, t_e), the functions $N \mapsto D'(t_2(C(N)))$ and $N \mapsto -D'(t_1(C(N)))$ are both increasing (decreasing), so C' is also increasing (decreasing) and C is therefore convex (concave). Yet, if one is convex and the other concave (as in Fig. 13(a)), it is still possible that the variations of one dominate the variations of the other, so that the overall function C is convex (as in Fig. 13(b)) or concave.

B Other proofs

Proof of Proposition 3. Differentiating Eq. (5) with respect to p leads to

$$\frac{\partial \mathrm{SC}}{\partial p}(S^P, p) = NC\left(\frac{p}{\bar{p}}N\right) + pNC'\left(\frac{p}{\bar{p}}N\right) - NC(N).$$

Provided that $p > 0$, this is of the same sign as

$$\frac{1}{p}\frac{\partial \mathrm{SC}}{\partial p}(S^P, p) = \frac{C\left(\frac{p}{\bar{p}}N\right) - C(N)}{p} + C'\left(\frac{p}{\bar{p}}N\right).$$

The convexity of C implies that both $\left(C(\frac{p}{\bar{p}}N) - C(N)\right)p^{-1}$ and $C'(\frac{p}{\bar{p}}N)$ are increasing, so their sum is increasing as well. Besides, $\frac{\partial \mathrm{SC}}{\partial p}(S^P, 0) = N(C(0) - C(N)) < 0$ and $\frac{\partial \mathrm{SC}}{\partial p}(S^P, \bar{p}) =$

$\bar{p}NC'(N) > 0$. Thus, for any given S^P, there exists a unique $p^* \in (0, \bar{p})$ minimizing $p \mapsto \mathrm{SC}(S^P, p)$ and $\mathrm{SC}(S^P, p)$ decreases with p on $p \in [0, p^*]$ and increases with p on $p \in [p^*, \bar{p}]$.

Let us now focus on the case $p = \frac{\bar{p}}{2}$. For this specific case, $\frac{C\left(\frac{p}{\bar{p}}N\right) - C(N)}{p}$ represents the negative of the average slope of the function $p \mapsto C\left(\frac{p}{\bar{p}}N\right)$ between the points $\frac{\bar{p}}{2}$ and $\bar{p}$. The convexity of C imposes that its absolute value is larger than $C'(N/2)$. Thus, $\frac{\partial \mathrm{SC}}{\partial p}\left(S^P, \frac{\bar{p}}{2}\right) \leq 0$, and therefore $\frac{\bar{p}}{2} \in (0, p^*]$.

Finally, the convexity of C also imposes that $C(N/2) \leq \frac{C(0)+C(N)}{2} = \frac{C(N)}{2}$, which means that $\mathrm{SC}(S^P, p^*) \leq \mathrm{SC}\left(S^P, \frac{\bar{p}}{2}\right) \leq \frac{\bar{p}}{2}\frac{C(N)}{2} + (1 - \frac{\bar{p}}{2})C(N) = \left(1 - \frac{\bar{p}}{4}\right)C(N)$. Thus, the maximum social cost reduction is of at least $\bar{p}/4$. □

Proof of Proposition 4. The congestion advantage A of priority users is continuous and decreases from $A(0) = C\left(\frac{N}{S}\right) - C(0)$ to $A(1) = 0$. Thus, the image of A is $\mathrm{Im}(A) = [0, A(0)]$. We are looking for $q^{EP} \in [0, 1]$ such that $F(A(q^{eq})) = q^{EP}$. This implies that q^{eq} belongs to the image of $\mathrm{Im}(A)$ under F. Since F is increasing, $q^{EP} \in [F(0), F(A(0))]$. Let then $F|_{\mathrm{Im}(A)}$ denote the restriction of F to $\mathrm{Im}(A)$. Assumption 4 ensures that the $F|_{\mathrm{Im}(A)}$ is continuous and strictly increasing, so we can define its inverse $F|_{\mathrm{Im}(A)}^{-1}$. The function $F|_{\mathrm{Im}(A)}^{-1}(q) - A(q)$ strictly increases on $[F(0), F(A(0))]$ from $-A(F(0)) \leq 0$ to $A(0) - A(F(A(0))) \geq 0$. Thus, $F|_{\mathrm{Im}(A)}^{-1}(q) - A(q)$ intersects 0 only once and there is a unique $q^{eq} \in [F(0), F(A(0))]$ such that $F|_{\mathrm{Im}(A)}^{-1}(q^{EP}) = A(q^{EP})$, i.e. $F|_{\mathrm{Im}(A)}(A(q^{EP})) = q^{EP}$. This solution is strictly greater than $F(0)$ if and only if $-A(F(0)) < 0$, i.e. $F(0) < \frac{gS^P}{S + (g-1)S^P}$. □