

The morning commute problem with temporary access restrictions for conventional and autonomous vehicles

Extended abstract submitted to the 6th symposium arranged by the European Association for Research in Transportation (hEART)

Raphaël Lamotte* André de Palma† Nikolas Geroliminis*

1 Introduction

The morning commute problem was introduced by Vickrey (1969). It considers a population of commuters who need to pass a single bottleneck of constant capacity. Commuters choose their departure time to minimize their generalized cost, which includes a travel time component and a schedule penalty for passing the bottleneck before or after their preferred arrival time. When the capacity constraint does not allow all users to pass the bottleneck at their preferred arrival time, a queue forms and forces users to choose some trade-off between low travel time and low schedule penalties.

As waiting times represent a deadweight loss, competition is a rather inefficient way to allocate the scarce bottleneck capacity. The classic remedy to this inefficiency is congestion pricing. Applied to the entire bottleneck capacity in a fine time-dependent manner, it can remove queuing altogether and bring the system to optimality (Vickrey, 1969), but it invariably leaves some low-income automobile users worse-off (Small, 1983). Applied on only a part of the road in the presence of severe congestion, it can reduce the social cost while ensuring a Pareto-improvement (Hall, 2016). Piece-wise constant tolls have also been proposed to facilitate implementation (Arnott et al., 1990). Despite these attractive variants, congestion pricing has rarely been implemented and researchers have turned to other alternatives that do not require any payment. One such alternative is the concept of *fast lane*, under which exclusive access to some part of the bottleneck capacity is granted to selected users during some time period (temporary HOV lanes for instance). Queues still form if the selected users need to compete between themselves. Yet, as the amount of queuing per passenger reflects the variability in schedule penalty in the allocated period, the queuing times can be significantly reduced if the allocated period covers a small range of schedule penalties. Fosgerau (2011) showed that with homogeneous users, such a scheme is equivalent to a coarse toll and produces a Pareto-improvement.

However, it is not clear yet to which extent “fast lanes” might be implemented in a fair and efficient manner in the real world. With homogeneous users, the optimum would involve creating as many small slots as users. Yet, such a solution can hardly be applied with heterogeneous users for several reasons.

*School of Architecture, Civil and Environmental Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland

†Ecole Normale Supérieure de Cachan, CES - Université Paris-Saclay, France

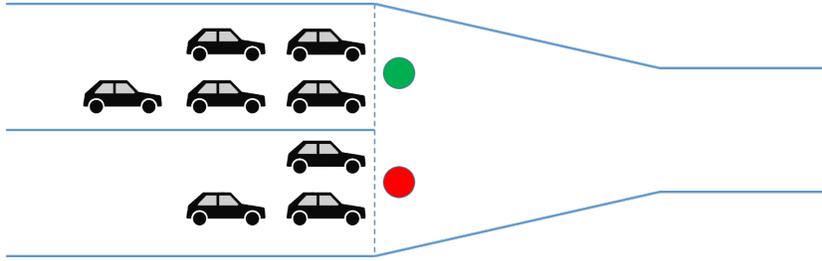


Figure 1: Illustration of a bottleneck divided between two populations

First, users’ scheduling preferences are generally unknown and ignoring them could be counterproductive and increase the social cost. Second, users with different access rights arriving at the same time should go to different queues (see Fig. 1 for an example with two populations), thereby limiting the number of user classes that can be considered in practice. Third, creating almost individual categories also raises equity and privacy-related issues commonly associated with non-anonymous pricing. In the present work, we consider only two categories based on vehicle type (autonomous and conventional vehicles) but introduce discrete intra-group heterogeneity in value of time and/or scheduling preferences. We show that in such a context, the socially optimal access restrictions apply to the entire bottleneck capacity but that heterogeneity in scheduling preferences may require defining multiple non-contiguous time periods for the same vehicle type.

2 Background

The choice of autonomous and conventional vehicles as our two vehicle classes is not simply trendy. Other criteria such as vehicle occupancy, fuel efficiency might also be selected to promote welfare-improving mobility choices. Yet, autonomous vehicles have another particularity: they are likely to make a very different usage of the bottleneck capacity. Some authors have argued that vehicle-to-vehicle communication would allow autonomous vehicles to follow each other much more closely, thereby doubling or tripling the effective capacity of roadways (Varaiya, 1993) and intersections (Lioris et al., 2017). If such predictions materialize even partially, autonomous vehicles would bring to the society benefits similar to those of public transit and might therefore deserve priority treatments at bottlenecks.

Consider a single bottleneck whose capacity can be divided into two in a time-dependent and continuous manner.¹ Denote $y(t)$ the proportion of the bottleneck capacity that is reserved to autonomous vehicles at time t . To avoid any ambiguity, we distinguish the nominal capacity S (defined as the maximum possible flow of conventional vehicles), and the effective capacity $Q(t)$ (which depends on time via the capacity split $y(t)$). We assume that the effective capacity varies linearly with the capacity split, i.e. that $Q(y(t)) = (1 - y(t))S + y(t)gS$. The positive coefficient g can be interpreted as the amount of capacity required per conventional vehicle, relatively to the amount required per autonomous vehicle, see Lamotte et al. (2017) for some order of magnitude.

In the following investigations, users are assumed to have the classic $\alpha - \beta - \gamma$ preferences, where α measures the cost of in-vehicle time, β the cost of an extra minute early at work and γ the cost of

¹One way to implement such a continuous split would be by metering parallel queues, similarly to what is done upstream of some tunnels and bridges.

an extra minute late (Arnott et al., 1990). More specifically, we will make the following assumption on the distribution of these parameters.

Assumption 1 (Discrete distribution of β/α). *Consider a population of heterogeneous commuters with $\alpha - \beta - \gamma$ preferences such that all users have the same ratio $\beta/\gamma = p$ and the same desired arrival time $t^* = 0$. The ratio β/α can only take m values r_1, r_2, \dots, r_m where $0 < r_1 \leq r_2 \leq \dots \leq r_m < 1$.*

Hereafter, we will denote n_i the size of the family having $\beta/\alpha = r_i$. Note that in the context of Assumption 1, all users of the same family have collinear vectors (α, β, γ) . It will be convenient to define α_i, β_i and γ_i as the mean values of α, β and γ among members of family i . Since all users are assumed to have the same ratio $\beta/\gamma = p$ and desired arrival time $t^* = 0$, we can also introduce the notion of *early equivalent* service time, which we define as $h(t) = t$ if $t < 0$ and $h(t) = -t/p$ if $t \geq 0$. In other words, $h(t)$ is the time before t^* that corresponds to the same schedule penalty as t .

To start, we consider the equilibrium within only one of the two classes of vehicles. Proposition 1 extends several well-known results concerning the equilibrium with heterogeneous users to the case with a time-dependent capacity.

Proposition 1. *Consider a population satisfying Assumption 1 that needs to pass a single bottleneck whose capacity at time t is given by a function $y(t)S$. The equilibrium satisfies the three following properties.*

1. *Users are all served in an interval $[t_0, t_1]$, where $t_1 - pt_0 = 0$ and such that $\int_{t_0}^{t_1} y(t)S = N$.*
2. *Let $(i, j) \in \{1, \dots, m\}^2, i < j$. Users of family j obtain service times that are more desirable than users of family i but experience a longer waiting time.*
3. *Denote t_i the early equivalent of the least desirable service time obtained by a member of family $i \in \{1, \dots, m\}$. The social cost is a continuous function of the set $\{t_1, \dots, t_m\}$ given by:*

$$SC(t_1, \dots, t_m) = \sum_{i=1}^m -n_i \beta_i t_i + \sum_{i=1}^{m-1} (t_{i+1} - t_i) r_i \left(\sum_{k=i+1}^m n_k \alpha_k \right). \quad (1)$$

Note that a very similar situation (without variations in the capacity) was considered in Arnott et al. (1988). The proof is almost the same. The originality of our statement however comes from the use of the times t_i to express the social cost, which greatly facilitates the derivations in the next section. For brevity and for reasons that will become clearer in the next section, we will refer to these times as *transition times*.

3 Optimal time-dependent access restrictions

This section investigates the socially optimal time-dependent access restrictions. We demonstrate that with the type of user heterogeneity considered in Assumption 1, a socially optimal time-dependent capacity split exists and although it is not necessarily unique, it must verify a few very stringent properties.

Proposition 2. *Consider a bottleneck that is shared between a population of autonomous vehicles (characterized by an effective capacity factor g) and a population of conventional vehicles. Assume that the two populations include heterogeneity as per Assumption 1. A socially optimal capacity split $y(t)$ exists.*

If the two populations include respectively l and m families, the proof of Proposition 2 makes use of the fact that the set of admissible transition times of autonomous families $t_{a,1}, t_{a,2}, \dots, t_{a,l}$ and conventional families $t_{c,1}, t_{c,2}, \dots, t_{c,m}$ form a closed and bounded subset of \mathbb{R}^{l+m} to apply the extreme value theorem and conclude. Admissible transition times are those that can be obtained at equilibrium by some function $y(t)$. Mathematically, the admissible tuples $(t_{a,1}, \dots, t_{a,l}, t_{c,1}, \dots, t_{c,m}) = (t_1, \dots, t_l, t_{l+1}, \dots, t_{l+m}) \in [t_0, 0]^{l+m}$ are such that $t_1 \leq \dots \leq t_l, t_{l+1} \leq \dots \leq t_{l+m}$, and

$$\forall i \in \{1, l+m\}, \int_{t_i}^{h(t_i)} y(u) S \, du \geq \sum_{j=1}^l \mathbf{1}_{[t_i, +\infty)}(t_j) \frac{n_j}{g} + \sum_{j=l+1}^{l+m} \mathbf{1}_{[t_i, +\infty)}(t_j) n_j, \quad (2)$$

where $\mathbf{1}_A(t)$ is the indicator function that is equal to 1 if t belongs to A and else is equal to 0.

Proposition 3. *Consider a bottleneck that is shared between a population of autonomous vehicles (characterized by an effective capacity factor g) and a population of conventional vehicles. Assume that the two populations include heterogeneity as per Assumption 1. A socially optimal capacity split $y(t)$ for the user equilibrium is such that each family i is served within two intervals $[t_i^-, t_i^+]$ and $[h(t_i^+), h(t_i^-)]$, and they benefit from the entire capacity during these intervals.*

Such strategies are commonly referred to in the control literature as *bang-bang* control. The proof of Proposition 3 starts from any strategy that is not bang-bang and shows that by increasing one of the transition times until Eq. (2) becomes an equality, the social cost is reduced. The only strategies such that all transition times satisfy (2) as an equality are actually bang-bang strategies.

Proposition 3 provides such a stringent characterization of the socially optimal capacity split that the number of candidate tuples for the social optimum is actually very limited. Since the sequence of families within each population and the durations they require are known, the only remaining degree of freedom is the sequence in which the two families are served. For two populations having respectively l and m families, there are essentially $\binom{l+m}{l}$ candidate tuples satisfying the necessary conditions for optimality presented in Proposition 3. If l and m are reasonably small, an exhaustive search can easily be conducted to identify the social optimum and otherwise, integer programming techniques should be used.

4 Policy implications

The choice of representative coefficients α_i, β_i and γ_i for each family has important consequences. In order to be consistent with the observed behaviours, the ratios between these coefficients should be estimated empirically. Yet, these ratios only impose that the chosen parameters are of the form $(k\alpha_i, k\beta_i, k\gamma_i)$, where $k > 0$ is a degree of freedom which should be set in agreement with sociopolitical objectives.

In a Kaldor-Hicks perspective, the infrastructure manager may choose to use the actual average values of α, β and γ observed within each family. Such a choice would result in additional weight being given to families who have high values of all three parameters. Although this is socially optimal insofar as rich individuals could hypothetically compensate poorer individuals for their loss, such a policy might not be desirable. Alternatively, the infrastructure manager may set the coefficients α, β and γ such that all families have the same value of α , of β , or using any other criterion.

For illustration purposes, we present the optimal configurations with only two families (flexible and inflexible) within each population and for three different scenarios, described in Table 1. The

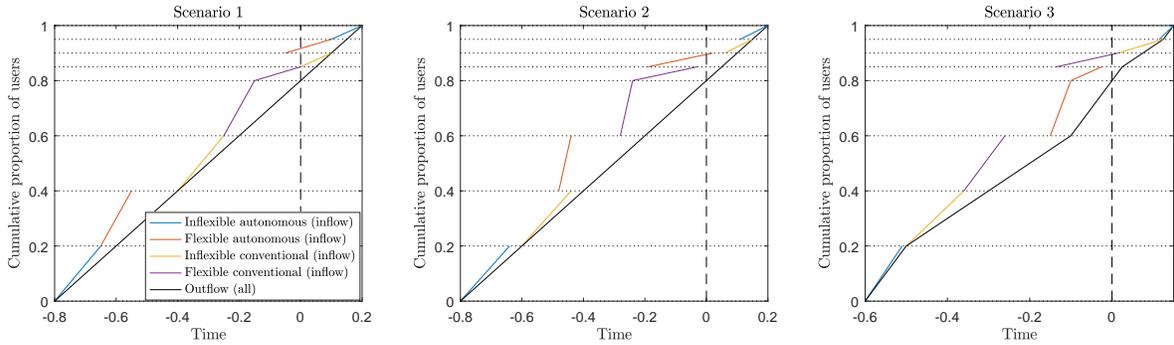


Figure 2: Dynamics with the socially optimal capacity allocation in three scenarios.

	Scenario 1 ($g = 1$)			Scenario 2 ($g = 1$)			Scenario 3 ($g = 2$)		
	α	β	β/γ	α	β	β/γ	α	β	β/γ
Flexible autonomous	2	0.5	0.25	1	0.2	0.25	1	0.1	0.25
Inflexible autonomous	1	0.5	0.25	1	0.8	0.25	1	0.5	0.25
Flexible conventional	2	0.5	0.25	1	0.2	0.25	1	0.3	0.25
Inflexible conventional	1	0.5	0.25	1	0.8	0.25	1	0.5	0.25

Table 1: Scheduling preferences of the different families in three scenarios.

four families have the same size in all three scenarios and flexible users have a ratio $\beta/\alpha < 0.5$ while inflexible users have a ratio $\beta/\alpha \geq 0.5$. As shown in Fig. 2, the three different configurations lead to three very different optimal ways of managing the capacity: either one type of vehicle travels in the wing, the other in the middle (Scenario 1), or the flexible and inflexible users of each type are separated by users from the other type (Scenario 2), or the flexible and inflexible users of only one type are separated by users from the other type, while the flexible and inflexible users of the other type travel in contiguous periods (Scenario 3).

References

- Arnott, R., De Palma, A., Lindsey, R., 1988. Schedule delay and departure time decisions with heterogeneous commuters. *Transportation Research Record* (1197), 56–67.
- Arnott, R., de Palma, A., Lindsey, R., 1990. Economics of a bottleneck. *Journal of urban Economics* 27 (1), 111–130.
- Fosgerau, M., 2011. How a fast lane may replace a congestion toll. *Transportation Research Part B: Methodological* 45 (6), 845–851.
- Hall, J. D., 2016. Pareto improvements from Lexus lanes: the effects of pricing a portion of the lanes on congested highways. Working paper, University of Toronto, accessed 2016-09-01.
URL http://individual.utoronto.ca/jhall/documents/PI_from_LL.pdf
- Lamotte, R., de Palma, A., Geroliminis, N., 2017. On the use of reservation-based autonomous vehicles for demand management. *Transportation Research Part B: Methodological* 99, 205–227.

Lioris, J., Pedarsani, R., Tascikaraoglu, F. Y., Varaiya, P., 2017. Platoons of connected vehicles can double throughput in urban roads. *Transportation Research Part C: Emerging Technologies* 77, 292–305.

Small, K. A., 1983. The incidence of congestion tolls on urban highways. *Journal of Urban Economics* 13 (1), 90–111.

Varaiya, P., 1993. Smart cars on smart roads: Problems of control. *IEEE Transactions on Automatic Control* 38, 195–207.

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