Karma Priority lanes for fair and efficient bottleneck congestion management

Conference Paper

Author(s):
Elokda, Ezzat; Cendese, Carlo; Zhang, Kenan; Censi, Andrea; Lygeros, John; Frazzoli, Emilio

Publication date:
2023

Permanent link:
https://doi.org/10.3929/ethz-b-000630557

Originally published in:
https://doi.org/10.1109/MED59994.2023.10185731

Funding acknowledgement:
180545 - NCCR Automation (phase I) (SNF)

Karma Priority lanes for fair and efficient bottleneck congestion management

Ezzat Elokda, Carlo Cendese, Kenan Zhang, Andrea Censi, John Lygeros and Emilio Frazzoli

Abstract— **A popular remedy for the morning commute bottleneck congestion is to split the highway capacity into a managed lane that is kept in free-flow and a general purpose lane that is subject to congestion. A classical theoretical result is that the more capacity is allocated to the managed lane the less the resulting congestion. However, existing approaches to restrict access to the managed lane are primarily monetary, e.g., tolls, which severely limits the public willingness to accept them due to equity concerns. Following up on recent work which introduces karma as a completely non-monetary credit used to control access to a so-called Karma Priority (KP) lane, we first review the strategic problem of the commuters which is modeled as a dynamic population game. We then numerically investigate the effect of varying the KP lane capacity. The karma scheme is equitable with respect to different income classes irrespective of the capacity split, meanwhile achieving near-optimal traffic reduction. Thus, managing a larger fraction of the bottleneck could be more socially feasible under a karma scheme than a monetary scheme.**

I. Introduction

For decades, traffic congestion has been causing tremendous social cost to major cities around the world. According to the INRIX 2019 Global Traffic Scorecard report, drivers lost an average of 99 hours a year in the U.S. and 115 hours a year in the U.K. due to congestion [1]. To manage rush hour traffic, a wide variety of tools have been proposed in the literature as well as implemented in practice. Among them, congestion pricing is the most widely known due to its theoretical efficiency [2]–[5]. However, the classical congestion pricing is often arguably politically and socially infeasible [6] as it tends to favor wealthier travelers [7]–[9]. For this reason, a growing attention has been drawn to alternative quantity- and credit-based approaches. The former directly limit the number of vehicles on the road, e.g., through license-plate rationing [10], [11] or highway reservation [12], [13], while the latter assign a limited number of travel credits/permits to road users which can be traded in a monetary market [14]–[16]. While these approaches are preferable to classical congestion pricing as they avoid a net financial flow from road users to authorities, they nevertheless fail to fundamentally address the equity issue. For instance, the license-plate rationing implemented in China has induced wealthy travelers to purchase additional vehicles [17]. Moreover, wealthy travelers can also take advantage of the credit-based schemes since they have a larger capacity to buy credits than others [18].

All authors are with ETH Zurich, 8092 Zurich, Switzerland. {elokdae, ccenedese, kenzhang, acensi, jlygeros, efrazzoli}@ethz.ch

Research supported by NCCR Automation, a National Centre of Competence in Research, funded by the Swiss National Science Foundation (grant number 180545).

During the morning commute bottleneck congestion [3], [19]–[21], commuters travel between a single origin-destination pair and arrive at a bottleneck forming a queue due to its limited capacity. A popular solution is to partially keep the bottleneck in free-flow by managing the access to some of its lanes, e.g., using High Occupancy Vehicle (HOV) lanes that are exclusive for carpoolers [22] or High Occupancy Toll (HOT) lanes that charge monetary tolls [23]. A well-known theoretical result is that if the managed lanes are successfully kept at their free-flow capacity, then the more lanes are managed the less the overall traffic congestion [23]. To this end, both HOV and HOT lanes face serious limitations. To ensure they are not wastefully under-filled, only a few lanes can be dedicated to HOV [24]. Due to the equity issue of congestion pricing and related monetary schemes, allocating more HOT lanes will likely lead to severe public dismay [9]. Moreover, it is practically difficult to determine the optimal toll charges as it requires authorities to have accurate measures of the commuters' private monetary Value of Time (VOT) [4], [25], [26].

To address these issues, in recent work [27], we propose a new kind of non-monetary mobility credits, called *karma*, that are used in an auction-like mechanism to fill a managed lane up to its free-flow capacity (hereafter referred to as the *Karma Priority (KP) lane*), while all commuters who fail to enter the managed lane use a General Purpose (GP) lane subject to congestion. We demonstrate that for a particular capacity split between the two lanes, the karma scheme gives rise to near-optimal traffic reduction without requiring any private information about the travelers, meanwhile addressing the equity issue because karma is completely decoupled from money. In this paper, our main contribution is in extending [27] to investigate the effect of increasing the relative capacity of the KP lane in order to maximize the benefit of the karma scheme. Our main findings is that near-optimal efficiency is robustly observed for most values of the capacity split, and significant traffic reduction can be achieved by dedicating up to 75% of the bottleneck capacity to the KP lane. We conjecture that due to the equity of the karma scheme, the public would be more willing to accept managing such high proportions of the bottleneck than under a monetary scheme. However, when the KP lane occupies the vast majority of the bottleneck (i.e., $\geq 80\%$), we observe equilibrium computation difficulties that point to potential limitations on how small the GP lane can be.

In the remainder of the paper, we first present a dynamic population game [28] model for the commuters' strategic problem under the presence of a KP lane, as first introduced in [27] (Section II). We then perform a numerical investigation of the effect of the KP lane capacity, demonstrating the achieved efficiency and equity of the karma scheme in comparison to an optimal monetary tolling scheme (Section III). We finally conclude with a discussion in Section IV.

A. Notation

Let $a, d \in D \subseteq \mathbb{N}$ and let $c \in C \subseteq \mathbb{R}^n$, then for a function $f : D \times C \rightarrow \mathbb{R}$, we distinguish discrete and continuous arguments through the notation $f[d](c)$. Alternatively, we write $f : C \rightarrow \mathbb{R}^{|D|}$ as the vector-valued function $f(c)$, with $f[d](c)$ denoting its d^{th} element. Similarly, $g[a \mid d](c)$ denotes the conditional probability of a given d and c. Specifically, $g[d^+ \mid d](c)$ denotes one-step transition probabilities for d. We denote by $p \in \Delta(D) := \left\{ \sigma \in \mathbb{R}_+^{|D|} \middle| \sum_{d \in D} \sigma[d] = 1 \right\}$ a probability distribution over the elements of D, with $p[d]$ denoting the probability of element d. Finally, when considering heterogeneous commuter types, we denote by x_τ a quantity associated to type τ.

II. Model

A. Description of CARMA *scheme*

We briefly recap CARMA, the karma-based bottleneck congestion management scheme first proposed in [27]. We consider N commuters that travel daily through a bottleneck with total capacity s [veh/min], which is split into a KP lane with capacity s_{fast} and a GP lane with capacity $s_{\text{slow}} = s - s_{\text{fast}}$. The commuters are heterogeneous in their *VOT process* (denoted by $\phi_\tau[u^+ \mid u]$ for commuters of type $\tau \in \Gamma = \{1, \dots, n_\tau\}$), that is, the process by which their daily VOT (denoted by $u \in \mathcal{U} = \{u_1, \dots, u_M\}$) changes. We discretize the feasible departure times into T intervals of length Δt [min]. Therefore, on every morning, commuters featuring state $[u, k]$ make a decision on their departure time $t \in \mathcal{T} = \{1, \dots, T\}$ and place a karma bid to enter the KP lane $b \in \mathcal{B}[k] = \{0, \dots, k\}$, where $k \in \mathbb{N}$ denotes their current budget in karma credits. Consequently, the highest $s_{\text{fast}}\,\Delta t$ bidders departing at t are allowed to enter the KP lane, while all others have to use the GP lane, as illustrated in Figure 1. Moreover, participation in the CARMA scheme is not mandatory; commuters not willing to participate can directly use the GP lane.

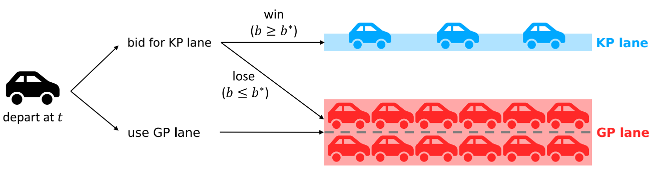


Fig. 1: Karma auction to enter KP lane.

B. Commuters' strategic model

We build upon the seminal Vickrey bottleneck congestion model [29] to incorporate it in the general karma dynamic population game model introduced in [30]. Let $g \in \Delta(\Gamma)$ be the distribution of VOT types in the population, $d \in \mathcal{D} = \{d \in \mathbb{R}^{|\Gamma| \times |\mathcal{U}| \times \infty} \mid \sum_{u,k} d_\tau[u, k] = g_\tau\}$ be the joint distribution of types and states, and $\pi \in \Pi$ be the policy, with $\pi_\tau[t, b \mid u, k] \in [0, 1]$ denoting the probability that commuters of type τ and state $[u, k]$ choose action $[t, b]$. The *social state* is $(d, \pi) \in \mathcal{D} \times \Pi$, which gives the distribution of commuter types and states as well as their actions, thereby providing a macroscopic description of the competitive landscape. The individual commuter faces a δ-discounted Markov decision process (MDP) that is coupled to others through the social state (d, π). In what follows, we specify the key elements of this MDP, those are, the immediate reward function $\zeta(d, \pi)$ and the karma transition function $\kappa(d, \pi)$.

1) Immediate reward function $\zeta[u, t, b](d, \pi)$: Following the classical bottleneck model, we define the immediate reward in two parts: queuing delay t^{q} and early or late schedule delay (t^{e} or t^{l}, respectively). Given the departure time and bid, how much delay each commuter endures depends on the outcome of the karma auction. Let $\psi[o \mid t, b](d, \pi)$ denote the probability of an ego commuter finally entering lane $o \in \{\text{fast}, \text{slow}\}$, given its choice of t, bid b and the other commuters' actions (function of the social state (d, π)). Then, the immediate reward can be written as

$$\zeta[u, t, b](d, \pi) = -u \sum_{o} \psi[o \mid t, b] \left(\alpha\, t^{\text{q}} + \beta\, t^{\text{e}} + \gamma\, t^{\text{l}} \right), \tag{1}$$

where t^{q}, t^{e} and t^{l} are given by (in (1) the dependency on the arguments $[t, o](d, \pi)$ is omitted for brevity):

$$t^{\text{q}}[t, o](d, \pi) = \begin{cases} 0, & o = \text{fast}, \\ \frac{q[t](d,\pi)}{s_{\text{slow}}}, & o = \text{slow}, \end{cases} \tag{2a}$$

$$t^{\text{e}}[t, o](d, \pi) = \max\{0,\ t^* - t - t^{\text{q}}[t, o](d, \pi)\}, \tag{2b}$$

$$t^{\text{l}}[t, o](d, \pi) = \max\{0,\ t + t^{\text{q}}[t, o](d, \pi) - t^*\}. \tag{2c}$$

In (1), the negation is to denote reward instead of cost, and $\beta < \alpha < \gamma$ give the sensitivity to the different delays. In (2a), $q[t](d, \pi)$ gives the queue length on the GP lane at time t, which is a function of the previous queue length $q[t-1](d, \pi)$ and the ratio of the GP lane departures to capacity. In (2b)–(2c), t^* is the commuters' desired arrival time.

To complete the definition of (1), we now derive $\psi[o \mid t, b](d, \pi)$. We define a threshold bid $b^*[t]$ such that

- if $b > b^*[t]$, the commuter enters the KP lane for sure, i.e., $o = \text{fast}$;
- if $b < b^*[t]$, the commuter enters the GP lane for sure, i.e., $o = \text{slow}$;
- if $b = b^*[t]$, the commuter ties with others and enters the KP lane via a random draw on the remaining capacity.

Let $\nu[t, b](d, \pi)$ be the mass of commuters departing at t and bidding b, i.e.,

$$\nu[t, b](d, \pi) = \sum_{\tau, u, k} d_\tau[u, k]\, \pi_\tau[t, b \mid u, k]. \tag{3}$$

Then, the threshold bid is given by

$$b^*[t](d, \pi) = \max \left\{ b \in \mathbb{N} \,\middle|\, \sum_{b' \geq b} \nu[t, b'] \geq \frac{s_{\text{fast}}}{N} \right\}. \tag{4}$$

Accordingly, the probability of entering the KP lane is derived as

$$\psi[o=\text{fast}\mid t,b](d,\pi)=\begin{cases}1, & b>b^*,\\ 0, & b<b^*,\\ \frac{s_{\text{fast}}/N-\sum_{b'>b^*}\nu[t,b']}{\nu[t,b]}, & b=b^*.\end{cases} \tag{5}$$

Note that $\psi[o=\text{fast}\mid t,b](d,\pi)$ is continuous in (d,π) except where $\nu[t,b](d,\pi)=0$. To guarantee the existence of a Stationary Nash Equilibrium (SNE) (see Section II-C), we approximate it with a function that is continuous everywhere, given by

$$\psi^\epsilon[o=\text{fast}\mid t,b](d,\pi)$$
$$=\begin{cases}1, & \sum_{b'>b}\nu[t,b']\le\frac{s_{\text{fast}}}{N}-\nu[t,b]-\epsilon,\\ 0, & \sum_{b'>b}\nu[t,b']\ge\frac{s_{\text{fast}}}{N},\\ \frac{s_{\text{fast}}/N-\sum_{b'>b}\nu[t,b']}{\nu[t,b]+\epsilon}, & \text{otherwise},\end{cases} \tag{6}$$

where $\epsilon>0$ is an arbitrarily small approximation parameter.

2) Karma transition function $\kappa[k^+\mid k,t,b](d,\pi)$: We consider a simple scheme where all commuters entering the KP lane pay their bids, and at the end of each day, the total payments are uniformly redistributed to all commuters in the system (refer to [27] for a treatment of different redistribution schemes). Let $p[b,o]$ be the karma payment made by a commuter who bids b, then we have

$$p[b,o]=\begin{cases}b, & o=\text{fast},\\ 0, & o=\text{slow}.\end{cases} \tag{7}$$

Accordingly, the *average payment* is computed by aggregating (7) over all commuters, i.e.,

$$\begin{aligned}\bar{p}(d,\pi)&=\sum_{t,b}\nu[t,b]\sum_{o}\psi^\epsilon[o\mid t,b]\,p[b,o]\\ &=\sum_{t,b}\nu[t,b]\,\psi^\epsilon[o=\text{fast}\mid t,b]\,b.\end{aligned} \tag{8}$$

To preserve the integer value of karma, $\lceil\bar{p}(d,\pi)\rceil$ is randomly distributed to a fraction of $f(d,\pi)=\bar{p}(d,\pi)-\lfloor\bar{p}(d,\pi)\rfloor$ of the commuters, and $\lfloor\bar{p}(d,\pi)\rfloor$ to the others. This yields the following karma transition probabilities, conditional on the outcome o:

$$\mathbb{P}[k^+\mid k,b,o](d,\pi)$$
$$=\begin{cases}f, & o=\text{fast and } k^+=k-b+\lceil\bar{p}\rceil,\\ 1-f, & o=\text{fast and } k^+=k-b+\lfloor\bar{p}\rfloor,\\ f, & o=\text{slow and } k^+=k+\lceil\bar{p}\rceil,\\ 1-f, & o=\text{slow and } k^+=k+\lfloor\bar{p}\rfloor,\\ 0, & \text{otherwise}.\end{cases} \tag{9}$$

Finally, we can construct the karma transition function as

$$\kappa[k^+\mid k,t,b](d,\pi)=\sum_{o}\psi^\epsilon[o\mid t,b]\,\mathbb{P}[k^+\mid k,b,o]. \tag{10}$$

C. Existence of Stationary Nash Equilibrium (SNE)

The Stationary Nash Equilibrium (SNE) is a social state (d^*,π^*) in which π^* is optimal for each individual commuter's MDP and additionally, d^* is stationary under the dynamics induced by π^*. Formally, it holds for all $\tau\in\Gamma$,

$$d^*_\tau=P_\tau(d^*,\pi^*)^\top\,d^*, \tag{11a}$$
$$\pi^*_\tau\in B_\tau(d^*,\pi^*), \tag{11b}$$

where $P_\tau(d,\pi)$ is the stochastic matrix for type τ's state and B_τ is the best response correspondence denoting the set of optimal policies for type τ's MDP, see [27], [28] for more details. As per [30], in order to guarantee the existence of a SNE, it must hold that $\zeta(d,\pi)$ and $\kappa(d,\pi)$ are continuous in (d,π), and additionally that the average amount of karma in the system is preserved in expectation. The former is straightforward to verify considering the continuous approximation (6), while the latter holds since all the karma payments are redistributed to the commuters, see [27] for a formal proof.

It follows that the karma economy for bottleneck congestion management is well-posed, i.e., a SNE is guaranteed to exist. The equilibrium can be computed using an evolutionary dynamics [31] inspired algorithm, as described in detail in [28], [30].

III. Effect of Karma Priority lane capacity

In this section, we compare the effect of the managed lane capacity between the KP scheme, a classical monetary HOT scheme, as well as the nominal benchmark in which no lane is managed (referred to as NOM). We first define the performance measures considered, then two numerical case studies that highlight the effect of varying the KP lane capacity are presented.

A. Performance measures

The performance measures can be divided into two categories: system level and user-type level. Specifically, we investigate the queuing delay and travel cost at the equilibrium traffic assignment. Table I summarizes the performance measures, where those associated with the benchmark (NOM) and the optimal tolling (HOT) are derived from the classical bottleneck model [21], [29], see [27] for more details. There, c^* is the equilibrium cost in the nominal case (normalized by the VOT multiplier u), $\bar{u}$ is the system level average VOT, $\bar{u}_\tau$ is the average VOT of type τ, and $\mathbb{P}_\tau[u]$ is the probability that commuters of type τ have VOT u (derived from the VOT process $\phi_\tau[u^+\mid u]$). The default values of the model parameters are reported in Table II.

B. Homogeneous commuters

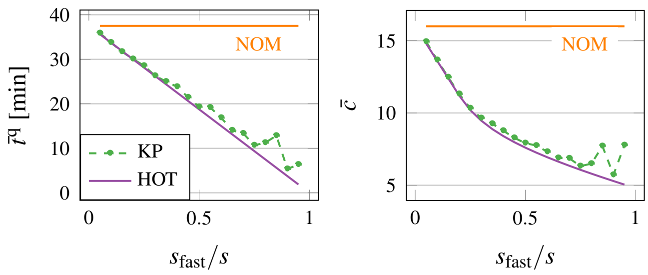
The purpose of this case study is to demonstrate the *efficiency* of the karma scheme with respect to HOT and NOM under the assumption that all commuters are homogeneous (and therefore there is no equity issue with HOT). We consider that all commuters have the same independent and identically distributed (i.i.d.) VOT process, namely, they have low VOT ($u_{\text{l}}=1$) 80% of the time and high VOT

TABLE I: Performance measures.

Name		**Benchmark** (“NOM”)	**Optimal tolling** (“HOT”)	**Karma scheme** (“KP”)*
System average queuing delay	$\bar{t}^{\mathrm{q}}$	$\frac{c^*}{2\alpha}$	$\frac{s_{\mathrm{slow}}}{s}\,\frac{c^*}{2\alpha}$	$\sum_{t,b}\nu[t,b]\sum_o \psi^\epsilon[o\|t,b]\; t^{\mathrm{q}}[t,o]$
System average travel cost	$\bar{c}$	$\bar{u}\; c^*$	$\sum_\tau g_\tau\; \bar{c}_\tau$	$-\sum_{\tau,u,k} d_\tau[u,k]\; R_\tau[u,k]$
Type average queuing delay	$\bar{t}^{\mathrm{q}}_\tau$	$\frac{c^*}{2\alpha}$	$\sum_u \mathbb{P}_\tau[u]\; \bar{t}^{\mathrm{q}}[u]$	$\frac{1}{g_\tau}\sum_{t,b}\nu_\tau[t,b]\sum_o \psi^\epsilon[o\|t,b]\; t^{\mathrm{q}}[t,o]$
Type average travel cost	$\bar{c}_\tau$	$\bar{u}_\tau\; c^*$	$\sum_u \mathbb{P}_\tau[u]\; \bar{c}[u]$	$-\frac{1}{g_\tau}\sum_{u,k} d_\tau[u,k]\; R_\tau[u,k]$
Type normalized travel cost	$\bar{c}^{\mathrm{n}}_\tau$	$\bar{c}_\tau/\bar{u}_\tau$		
*all measures are computed at the SNE.				

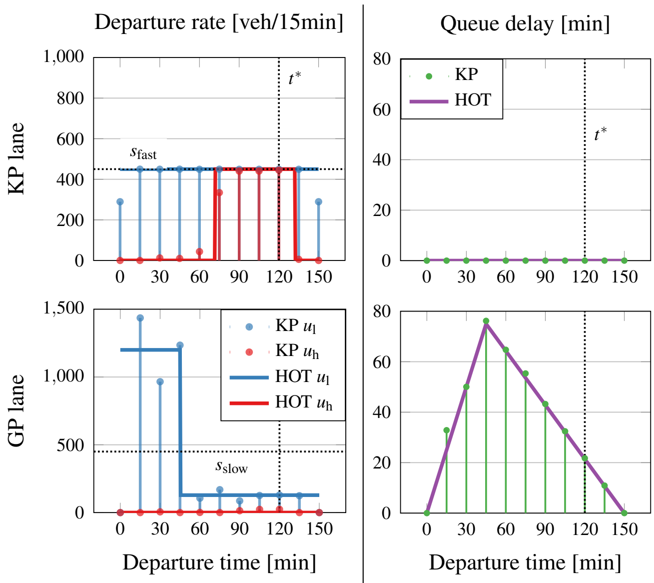
TABLE II: Default values of model parameters.

Name	**Notation**	**Unit**	**Value**
Number of commuters	N		9000
Bottleneck capacity	s	veh/min	60
Length of discrete time step	Δt	min	15
Normalized VOT		cost/hour*	
- queuing delay	α		6.4
- early arrival	β		4
- late arrival	γ		16
Desired arrival time	t^*	min	120
Discount factor	δ		0.99
Parameter for model continuity	ϵ		10^{-4}
Average karma per commuter	$\bar{k}$		10
*in HOT the unit is \$/hour.			



(a) (b)

Fig. 2: System average queuing delay (a) and average travel cost (b) as a function of the fraction of the managed lane s_{fast}/s for homogeneous commuters. The values for NOM are computed at $s_{\mathrm{fast}}/s = 0$ and displayed for reference.



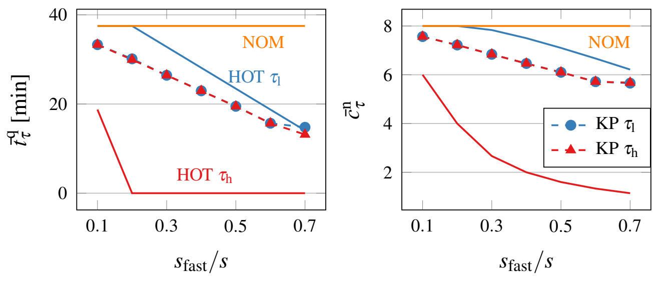
(a) (b)

Fig. 3: Departure rate (a) and queueing delay (b) in the KP and GP lane as a function of the departure time t for homogeneous commuters and $s_{\mathrm{fast}}/s = 0.5$.

($u_{\mathrm{h}} = 6$) 20% of the time. This is represented by a Markov chain with stochastic matrix $\phi = \begin{pmatrix} 0.8 & 0.2 \\ 0.8 & 0.2 \end{pmatrix}$.

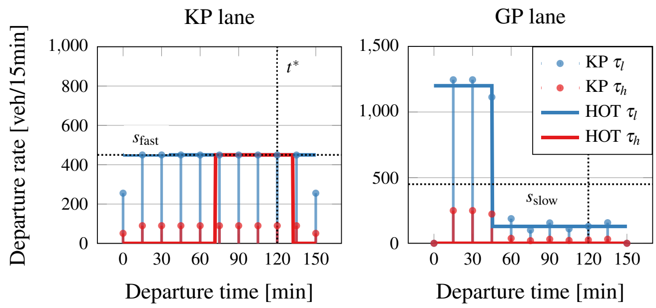
Figure 2 shows the system average queuing delay (Figure 2a) and average travel cost (Figure 2b) as a function of the fraction of the managed lane to total capacity $s_{\mathrm{fast}}/s \in \{0.05, \ldots, 0.95\}$. We observe the classical theoretical result that under the optimal monetary tolling (HOT), both performance measures monotonically decrease as the managed lane capacity increases; hence it is most efficient to toll as much of the bottleneck capacity as possible. The average queuing delay decreases linearly, while the marginal decrease in the average travel cost is higher at small s_{fast}, up to roughly when the managed lane can accommodate all the high VOT commuters (i.e., $s_{\mathrm{fast}}/s = 0.2$ in this case). Remarkably, for most considered values of s_{fast}/s, the karma scheme (KP) closely follows the optimal queuing delay and travel cost reduction of HOT. For example, at $s_{\mathrm{fast}}/s = 0.75$, KP (HOT) achieves an average queuing delay reduction of 71.5% (75%) and an average travel cost reduction of 60.5% (62.1%) with respect to NOM, hence there is insignificant loss in efficiency by restricting access to the managed lane using karma instead of money. However, when the KP lane occupies the vast majority of the bottleneck (i.e., $s_{\mathrm{fast}}/s \geq 80\%$), we observe fluctuations in the performance of KP which we attribute to convergence difficulties of the SNE algorithm in this regime. This points to potential limitations on how small the GP lane can be made.

To shed light on the efficiency of KP, Figure 3 shows the departure rates (Figure 3a) and queuing delays (Figure 3b) in both lanes at the SNE of KP for the bottleneck capacity split of $s_{\mathrm{fast}}/s = 0.5$, compared to the equilibrium of HOT.



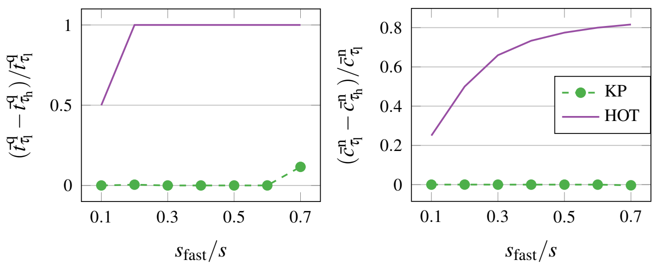
(a) (b)

Fig. 4: User-type average queuing delay (a) and normalized travel cost (b) as a function of the fraction of the managed lane s_{fast}/s for heterogeneous commuters. The values for NOM are computed at $s_{\text{fast}}/s = 0$ and displayed for reference.



(a) (b)

Fig. 6: Departure rate in the KP lane (a) and GP lane (b) as a function of the departure time t for heterogeneous commuters and $s_{\text{fast}}/s = 0.5$.



(a) (b)

Fig. 5: Relative advantage of the high income group τ_{h} compared to the low income group τ_{l} in terms of the average queuing delay (a) and normalized travel cost (b) as a function of the fraction of the managed lane s_{fast}/s.

As expected, the KP lane does not experience queuing delays, whereas the queuing delays in the GP lane follow the characteristic shape of the classical bottleneck model. Similarly to HOT, the least costly departures, viz. those closest to t^* on the KP lane, are dominated by the high VOT commuters. Thus the KP lane is allocated efficiently. In fact, the future-sighted commuters prefer to yield the KP lane when having low VOT to save karma for future high VOT events.

C. Heterogeneous commuters

This case study demonstrates the *equity* of KP in comparison to HOT and how the capacity of the KP lane exacerbates this gap. We consider two groups of commuters: τ_{l} has VOT $u = 1$ all the time, and τ_{h} has VOT $u = 6$ all the time, i.e., $\mathbb{P}_{\tau_{\text{l}}}[u = 1] = 1$ and $\mathbb{P}_{\tau_{\text{h}}}[u = 6] = 1$. Empirical evidence has shown that the individual (monetary) VOT is highly correlated with the income level [32]. Hence, we may consider commuters in group τ_{h} as *wealthier* than those in group τ_{l}. The low income group τ_{l} occupies 80% of the population and the high income group τ_{h} the remaining 20%.

Figure 4 shows the average queuing delay (Figure 4a) and the normalized travel cost[1] (Figure 4b) of each income group as a function of the fraction of the managed lane to total capacity $s_{\text{fast}}/s \in \{0.1, \dots, 0.7\}$[2]. We observe a stark contrast in the equity of KP and HOT. Under HOT the low income group is significantly disadvantaged, as is further demonstrated in Figure 5 showing the relative advantage of the high income group compared to the low income group for the same performance measures in Figure 4. When the managed lane can accommodate all the high income commuters, i.e., $s_{\text{fast}}/s \geq g_{\tau_{\text{h}}} = 0.2$, those commuters do not incur any queuing delays which are instead entirely experienced by the low income commuters (see Figures 4a and 5a). Moreover, the relative disparity in the normalized travel cost is monotonically increasing with the fraction of the managed lane (see Figure 5b); at $s_{\text{fast}}/s = 0.7$ the high income commuters incur 81.2% less travel cost than the low income commuters.

In contrast, KP is invariant to income and does not discriminate against the low income group: both groups experience the same average queuing delay and travel cost reduction irrespective of the bottleneck capacity split. Figure 6 sheds light on this by showing the departure rates in both lanes when $s_{\text{fast}}/s = 0.5$. Under HOT, the most desired times on the managed lane are occupied by the high income group (see Figure 6a), whereas under KP, the managed lane capacity is split between the two groups in the same ratio as their proportions in the population. This is because under KP, both groups essentially face the same optimization problem, up to a constant difference in the scaling of the VOT.

IV. Conclusions

We consider the morning commute bottleneck congestion problem and introduce the concept of Karma Priority (KP) lanes to control access to part of the bottleneck capacity. KP lanes are similar in principle to High Occupancy Toll (HOT) lanes, which are well known to lead to significant traffic reduction if the toll charges are set optimally. The

[1]The normalization is with respect to the average VOT/income level of each group and brings the costs of both groups on the same scale.

[2]The SNE algorithm did not converge for $s_{\text{fast}}/s \geq 0.8$.

distinguishing feature of our approach is that we use a completely non-monetary karma economy to manage access to the KP lanes. Commuters participate in karma auctions to enter the KP lane, and the total karma payments of those entering the lane get redistributed to all commuters at the end of each day.

Through numerical analysis we demonstrate that our proposed karma scheme achieves near-optimal traffic reduction, both in terms of the average travel time and the average user travel cost. Thus there is negligible loss in efficiency by adapting a karma scheme instead of a monetary one. Importantly, however, there is a significant gain in the equity with respect to different income classes. While a monetary scheme severely discriminates against low income groups, the karma scheme is invariant to income and thus equitable. Therefore, we conjecture that the public would be willing to dedicate more of the bottleneck capacity as KP lanes, leading to more traffic reduction.

Future work includes performing an in-depth analysis of the limiting regime when the vast majority of the bottleneck capacity is used as a KP lane. Our macroscopic model does not capture practical limitations in how small the General Purpose (GP) lane can be. We suspect that this could be related to our observed equilibrium computation instabilities for very high KP lane capacity. Moreover, it is interesting to extend our model to account for multi-occupancy vehicles and investigate karma policies that incentivize carpooling, e.g., by combining the karma of carpoolers.

References

[1] INRIX, "2019 global traffic scorecard," 2020.

[2] A. C. Pigou, *The economics of welfare*. Macmillan and Company, London, 1920.

[3] W. S. Vickrey, "Congestion theory and transport investment," *The American Economic Review*, vol. 59, no. 2, pp. 251–260, 1969.

[4] A. De Palma and R. Lindsey, "Traffic congestion pricing methodologies and technologies," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1377–1399, 2011.

[5] C. Cenedese, P. Stokkink, N. Geroliminis, and J. Lygeros, "Incentive-based electric vehicle charging for managing bottleneck congestion," *European Journal of Control*, p. 100697, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0947358022000905

[6] D. K. Brands, E. T. Verhoef, J. Knockaert, and P. R. Koster, "Tradable permits to manage urban mobility: market design and experimental implementation," *Transportation Research Part A: Policy and Practice*, vol. 137, pp. 34–46, 2020.

[7] R. Arnott, A. De Palma, and R. Lindsey, "The welfare effects of congestion tolls with heterogeneous commuters," *Journal of Transport Economics and Policy*, pp. 139–161, 1994.

[8] A. W. Evans, "Road congestion pricing: when is it a good policy?" *Journal of transport economics and policy*, pp. 213–243, 1992.

[9] B. D. Taylor and R. Kalauskas, "Addressing equity in political debates over road pricing: Lessons from recent projects," *Transportation research record*, vol. 2187, no. 1, pp. 44–52, 2010.

[10] X. Wang, H. Yang, and D. Han, "Traffic rationing and short-term and long-term equilibrium," *Transportation research record*, vol. 2196, no. 1, pp. 131–141, 2010.

[11] D. Han, H. Yang, and X. Wang, "Efficiency of the plate-number-based traffic rationing in general networks," *Transportation Research Part E: Logistics and Transportation Review*, vol. 46, no. 6, pp. 1095–1110, 2010.

[12] D. Teodorovic, P. Edara, and C. E. Via, "Highway space inventory control system," in *16th International Symposium on Transportation and Traffic Theory*, 2005.

[13] P. Edara and D. Teodorović, "Model of an advance-booking system for highway trips," *Transportation Research Part C: Emerging Technologies*, vol. 16, no. 1, pp. 36–53, 2008.

[14] E. Verhoef, P. Nijkamp, and P. Rietveld, "Tradeable permits: their potential in the regulation of road transport externalities," *Environment and Planning B: Planning and Design*, vol. 24, no. 4, pp. 527–548, 1997.

[15] D. Teodorović, K. Triantis, P. Edara, Y. Zhao, and S. Mladenović, "Auction-based congestion pricing," *Transportation Planning and Technology*, vol. 31, no. 4, pp. 399–416, 2008.

[16] H. Yang and X. Wang, "Managing network mobility with tradable credits," *Transportation Research Part B: Methodological*, vol. 45, no. 3, pp. 580–594, 2011.

[17] Y. Nie, "Why is license plate rationing not a good transport policy?" *Transportmetrica A: Transport Science*, vol. 13, no. 1, pp. 1–23, 2017.

[18] F. Xiao, Z. S. Qian, and H. M. Zhang, "Managing bottleneck congestion with tradable credits," *Transportation Research Part B: Methodological*, vol. 56, pp. 1–14, 2013.

[19] C. Hendrickson and G. Kocur, "Schedule delay and departure time decisions in a deterministic model," *Transportation science*, vol. 15, no. 1, pp. 62–77, 1981.

[20] C. F. Daganzo, "The uniqueness of a time-dependent equilibrium distribution of arrivals at a single bottleneck," *Transportation science*, vol. 19, no. 1, pp. 29–37, 1985.

[21] R. Arnott, A. De Palma, and R. Lindsey, "Economics of a bottleneck," *Journal of urban economics*, vol. 27, no. 1, pp. 111–130, 1990.

[22] J. Dahlgren, "High occupancy vehicle lanes: Not always more effective than general purpose lanes," *Transportation Research Part A: Policy and Practice*, vol. 32, no. 2, pp. 99–114, 1998.

[23] ——, "High-occupancy/toll lanes: where should they be implemented?" *Transportation Research Part A: Policy and Practice*, vol. 36, no. 3, pp. 239–255, 2002.

[24] J.-P. Wang, H.-J. Huang, and X. J. Ban, "Optimal capacity allocation for high occupancy vehicle (hov) lane in morning commute," *Physica A: Statistical Mechanics and its Applications*, vol. 524, pp. 354–361, 2019.

[25] H. Yang, Q. Meng, and D.-H. Lee, "Trial-and-error implementation of marginal-cost pricing on networks in the absence of demand functions," *Transportation Research Part B: Methodological*, vol. 38, no. 6, pp. 477–493, 2004.

[26] R. Lindsey, "Do economists reach a conclusion?" *Econ Journal Watch*, vol. 3, no. 2, pp. 292–379, 2006.

[27] E. Elokda, C. Cenedese, K. Zhang, J. Lygeros, and F. Dörfler, "Carma: Fair and efficient bottleneck congestion management with karma," *arXiv preprint arXiv:2208.07113*, 2022.

[28] E. Elokda, A. Censi, and S. Bolognani, "Dynamic population games," *arXiv preprint arXiv:2104.14662*, 2021.

[29] R. Arnott, A. De Palma, and R. Lindsey, "A structural model of peak-period congestion: A traffic bottleneck with elastic demand," *The American Economic Review*, pp. 161–179, 1993.

[30] E. Elokda, S. Bolognani, A. Censi, F. Dörfler, and E. Frazzoli, "A self-contained karma economy for the dynamic allocation of common resources," *arXiv preprint arXiv:2207.00495*, 2022.

[31] W. H. Sandholm, *Population games and evolutionary dynamics*. MIT Press, 2010.

[32] M. Börjesson, M. Fosgerau, and S. Algers, "On the income elasticity of the value of travel time," *Transportation Research Part A: Policy and Practice*, vol. 46, no. 2, pp. 368–377, 2012.