Traffic priority mechanisms

Roman Zakharenko[1]

National Research University Higher School of Economics, Moscow, Russia. Mailing address: office 3433, Shabolovka 26, Moscow 119049 Russia. Email: rzakharenko@hse.ru. Web: www.rzak.ru. Phone: +7-495-772-9590, ext.026081

Abstract

In the future world of self-driving vehicles, intersections will be managed by computers that send individual commands to each passing vehicle. This paper proposes to make traffic priority contingent on self-reported value of time of vehicle occupants. A model of two merging roads with stochastic traffic is developed. Algorithms for calculation of optimal exit sequences, accounting for time value heterogeneity, are characterized. Welfare costs of limited planning horizon are assessed. Incentive compatible scheme of payment for priority is calculated. The winners and losers of the proposed mechanism are described. Optimality of traffic volume and composition under optimal exit regulation is established.

Keywords: Traffic congestion, Automated traffic management, Private value of time, Mechanism design, Priority queue

JEL codes: C63, D82, R41, R48

1. Introduction

The introduction of self-driving vehicles will also change other transport infrastructure. In particular, it is expected that modern traffic lights at road intersections will be replaced by computers that send individual commands to every passing vehicle. Because of instant reaction of computerized vehicles, it is expected that intersection throughput capacity will dramatically increase.

April 24, 2018

The prospect of making traffic commands vehicle-specific creates an opportunity to improve traffic management even further by considering the problem from the economics perspective. Different travelers may have different values of time that they spend in traffic delays. This paper proposes to optimize traffic priority in accordance with self-reported value of time by occupants of each vehicle passing through an intersection. Those who report higher time value get priority, in exchange for a toll payment collected by the regulating authority. Essentially, every traveler is allowed to buy (a higher probability of) a green traffic light at each intersection they pass.

For the purposes of mathematical tractability, I consider the most simple type of transport infrastructure that causes congestion, a merger of two one-way roads into a single road. On each of two entry roads, vehicles arrive sequentially and form a queue near the intersection. The authority then has to decide, at any moment of time, which of the two queues should be given priority in exiting the intersection. Because travelers' time value is their private information, principles of mechanism design have to be applied to determine the optimal payment for passing through the intersection, as a function of reported time value.

This paper pursues answers to several questions. First, what is the optimal exit sequence of the vehicles queued at the intersection? Second, if a queue is very long, how much welfare is lost if the authority considers only part of the queue to speed up computation of optimal exit sequence? Third, what is the optimal payment scheme to induce truthful reporting of the time values? Fourth, assuming all payment proceeds are rebated to the population as lump-sum transfers, are there any groups whose welfare is reduced by optimal regulation? Finally, if decisions to travel through the intersection are endogenous, can the first-best level of welfare be achieved?

Wherever possible, the paper addresses these questions analytically, making the results general.

All conclusions of this paper are applicable not only to automobiles, but also to any other

transport system with short intervals between vehicles. For example, futuristic personal rapid transport systems such as SkyTran can benefit from the ideas of this paper.

This paper relates multiple, previously unrelated, strands of literature. The earliest relevant contributions are probability theory studies, such as Tanner (1962) and Cowan (1979). Both study a merger of two roads, very similar to that of this paper, with randomly arriving vehicles. These studies focus on questions such as the expected delay and the size distribution of gaps between vehicles induced by the merging process. Heterogeneity of travelers' time values is not discussed.

On the side of traffic engineering, many studies analyze how automation will affect intersection throughput. Kamal et al. (2015) simulate such an intersection and conclude that its capacity may be increased by nearly 100% relative to modern intersections. Milanés et al. (2011) use actual automobiles with automated and coordinated lateral controls to conduct a physical simulation of merging traffic, with encouraging results for increasing road capacity. Wang et al. (2007) argue that advance planning of vehicle merge timing and sequence may have a further dramatic increase in throughput. Milakis et al. (2017) provide further references on the topic. In this strand of literature, heterogeneity of travelers is out of question, too.

As this paper introduces the notion of priority into traffic queues, it is somewhat related to priority queues studied in computer science. But there are also substantial differences, too. First, in computer science, queued elements are tasks that can be served in an arbitrary sequence. In traffic studies, queued elements are vehicles with physical dimensions (size, mass) that cannot be instantly reshuffled. Most such studies, including this one, assume that vehicles from a given entry road cannot pass each other and must exit in order of arrival. The regulating authority can only decide which of entry roads gets a green light at each moment of time. Second, the queued elements in the computer science always arrive unexpectedly. In traffic, because information about vehicles can travel faster than vehicles

themselves, such arrivals can be foreseen. Section 3.2 provides an example in which the optimal exit sequence is affected by a vehicle that did not yet join the queue, and develops an algorithm for calculation of the optimal exit sequence in such settings.

A number of computer science studies, for example Nisan and Ronen (1999) and Feigenbaum and Shenker (2002), propose to account for self-interest of participating agents when developing optimization algorithms, e.g. for optimal use of computer power on a network of computers. These theories are however focused on self-interested *providers* of capacity, while the current paper assumes self-interested *users* (i.e. travelers), and the capacity (i.e. the road infrastructure) is assumed to be provided by a benevolent government.

Finally, due to continuous nature of vehicle arrivals to an intersection, this paper is related to the mechanism design literature with dynamic populations. Bergemann and Said (2011), Said (2012), Parkes and Singh (2004) are the relevant references in this field.

2. A model of an intersection

In a future world of self-driving vehicles, consider a merger of two one-way roads into one. We will label the two entry roads as *Left*, denoted L, and *Right*, denoted R. All vehicles are assumed to have a standardized length of one unit, including the minimum safety headway between the vehicles at the cruise speed. Time is continuous, and is normalized so that the cruise speed of a vehicle is equal to one.

Vehicles arrive randomly at the intersection from both entry roads. Vehicles may arrive in *bunches*, i.e. with minimum unitary intervals between vehicles, or with *gaps*, i.e. with intervals exceeding unity. At a given moment of time, denote the value of a unit of time for the users of the i-th vehicle (counting from the intersection) on road $S \in \{L, R\}$ by $v_i^S \geq 0$. This time value is private information and is drawn from a commonly known distribution. The distribution is assumed to be the same for both Left and Right roads, and the time value draws are independent across any two vehicles. At a given moment of time, there are

4

$n_S \geq 0$ vehicles on road $S \in \{L, R\}$. By vehicle *layout* we denote all information about vehicles known to approach the intersection at a given point in time, i.e. from which road they enter, their distance to the intersection, and their time values.

The intersection is managed by a computerized *authority* which communicates with passing vehicles by radio. Some distance before the intersection, the vehicles automatically report the time value of their users to the authority and submit a toll payment which is a function of the reported time value.[2] The authority then calculates and announces the exact time when each vehicle should clear the intersection. All vehicles must clear the intersection at the cruise speed, with the minimum interval (of both time and space) equal to one.[3] The sequence in which vehicles clear the intersection is labeled the *exit sequence*; the authority chooses the exit sequence that minimizes the welfare loss due to congestion. The exit sequence may be updated as new vehicles arrive.

3. Optimal exit sequence

Throughout the paper, optimal exit sequences will be calculated only for the approaching vehicles that have already reported their presence to the authority, and disregarding potential additional arrivals. This allows us not to calculate the expected future states of the system, thereby decreasing the dimensionality of the problem. Taking expected future arrivals into account vastly increases computational difficulty, while the gains in terms of social welfare are questionable. The latter is especially true when approaching vehicles report themselves sufficiently early (i.e. those who still did not report themselves are faraway), and when the

---

[2]In a more complex model with a network of roads and multiple intersections, the value of time needs to be reported only once for each trip, and payment depends not only on the time value but also on the route traveled.

[3]There is a large literature on "gap acceptance" which highlights the fact that driver psychology may greatly affect exit speed and intervals when there is merging traffic, e.g. Pollatschek et al. (2002). This paper assumes that the problem will be eliminated by automation of driving, so the vehicles exit at maximum speed with minimum intervals.

time value distribution is the same for vehicles on both entry roads. Welfare gains of greater reporting and planning horizon are discussed in section 3.3.

3.1. Queues without gaps

Consider first a vehicle layout in which there is a bunch of n_L vehicles on the Left road and a bunch of n_R vehicles on the Right road. No vehicles have gaps ahead of them, and therefore the first vehicles from each entry road are able to clear the intersection immediately if given priority. The Left vehicles are labeled $L_1, \ldots, L_{n_L}$, while the Right vehicles are labeled $R_1, \ldots, R_{n_R}$. The reported values of vehicles on road $S \in \{L, R\}$ are labeled $\{v_1^S, \ldots, v_{n_S}^S\}$. No additional vehicles are present or expected on either entry road. In this setting, vehicles always exit as a single bunch of $n_L + n_R$ vehicles, without gaps between them. Thus, the total exit time is $n_L + n_R$ regardless of the chosen exit sequence of vehicles; the exit time of a given vehicle is uniquely determined by its position in the exit sequence.

The total number of possible exit sequences is $C_{n_L+n_R}^{n_L} = \frac{(n_L+n_R)!}{n_L!n_R!}$. Verifying all of them, however, is not necessary due to a nice property of queues without gaps: the delay of a vehicle is uniquely determined by its location in the exit sequence, and does not depend on location of other vehicles in that sequence. Therefore, when considering reshuffling some part of the exit sequence, the authority should compare only the time values of the vehicles being reshuffled, and can ignore vehicles whose position in the sequence does not change. This property enables us to specify a fast algorithm for determining the optimal exit sequence, with computational time proportional to $n_L + n_R$. The algorithm makes use of an "average-down" partitioning of all vehicles on each road, described below.

3.1.1. Description of optimal exit sequence

Average-down partitioning of vehicles on a road. For each road $S \in \{L, R\}$, define $a^S(n_1, n_2) \equiv \frac{\sum_{i=n_1}^{n_2} v_i^S}{n_2 - n_1 + 1}$ as the average time value of all vehicles on road S, positioned between n_1 and n_2.

First element of partitioning Define $g_1^S \equiv \arg\max_{g=1...n_S} a^S(1, g)$ as the index of the vehicle g that maximizes the average time value of vehicles from 1 to g. If multiple g's maximize $a^S(1, g)$, then g_1^S is the smallest (i.e. earliest) such g. Define by $G_1^S = \{S_1, \ldots, S_{g_1^S}\}$ the bunch of vehicles from S_1 to $S_{g_1^S}$. By construction, this is the bunch of vehicles with the highest average value of time on road S, among all bunches that include the first vehicle S_1. G_1^S is the first element of the vehicle partitioning on road S.

Further elements of partitioning For any $k = 1, 2, 3, \ldots$, if there are any vehicles on road S that are *not* included into previously defined bunches $\{G_1^S, \ldots, G_k^S\}$, define $g_{k+1}^S \equiv \arg\max_{g=g_k^S+1,\ldots,n_S} a^S(g_k^S + 1, g)$ as the index of the vehicle g that maximizes the average time value of vehicles from $g_k^S + 1$ to g. In case of maximization ties, g_{k+1}^S is the smallest (earliest) possible candidate. Define by $G_{k+1}^S = \{S_{g_k^S+1}, \ldots, S_{g_{k+1}^S}\}$ the bunch of vehicles that were not included into $G_1^S, \ldots, G_k^S$ and that maximize the average time value among the remainder.

This completes the partitioning algorithm. We refer to the elements G_k^S of the partitioning as *sets*, to differentiate them from arbitrary vehicle bunches. By construction, the average time value of vehicles in set G_{k+1}^S is weakly lower than that in G_k^S, which motivates the "average-down" nickname of the partitioning. The maximum possible number of sets k on road S in equal to the number of vehicles on that road; such number of sets is realized when vehicle time values are weakly decreasing.

The following lemma is useful in subsequent analysis.

Lemma 1. *For each set G_j^S in the average-down partitioning that has more than one element, for each vehicle i except the last one g_j^S in that set, we have that $a^S(g_{j-1}^S + 1, i) < a^S(g_{j-1}^S + 1, g_j^S) < a^S(i + 1, g_j^S)$. In words, if a set is partitioned into two bunches, the earlier bunch has strictly lower average time value.*

Proof. By construction, the last vehicle of the set g_j^S was chosen to maximize the average

$a^S(g_{j-1}^S + 1, g)$, thus

$$a^S(g_{j-1}^S + 1, i) < a^S(g_{j-1}^S + 1, g_j^S). \tag{1}$$

The average time value of the set is also the weighted average of earlier and of later vehicles within the set, thus

$$a^S(g_{j-1}^S + 1, g_j^S) < a^S(i + 1, g_j^S). \tag{2}$$

Inequalities (1,2) together prove the Lemma. ∎

We now return to the problem with two merging roads and characterize the optimal exit sequence. Applying the average-down partitioning to vehicles on each of the roads results in sets $\{G_1^L, \ldots, G_{k_L}^L\}$ and $\{G_1^R, \ldots, G_{k_R}^R\}$, respectively. Denote by $|G_i^S| \equiv a^S(g_{i-1}^S + 1, g_i^S)$ the average time value of vehicles in set G_i^S.

Theorem 1. *In the optimal exit sequence, vehicles exit in sets $\{G_1^L, \ldots, G_{k_L}^L, G_1^R, \ldots, G_{k_R}^R\}$, so that the sets are sorted in order of decreasing average time values.*

Proof. Consider the optimal exit sequence, i.e. the one that minimizes the total welfare lost in delays. Partition the exit sequence into bunches $\{B_1, \ldots, B_k\}$, such that each bunch consists of vehicles from the same road, and each bunch is followed (except the last one) and preceded (except the first one) by vehicles from the other road. Without loss of generality, assume that B_1 consists of vehicles from the Left; then, every odd-numbered B_i is from Left, while every even-numbered B_i is from Right. Apply the average-down partitioning to each bunch $B_i, i = 1, \ldots, k$ so it becomes subdivided into sets $B_{i,1}, \ldots, B_{i,m_i}$ with weakly decreasing average time values, $|B_{i,1}| \geq |B_{i,2}| \geq \cdots \geq |B_{i,m_i}|$.

Suppose there exist two adjacent sets from opposite roads such that the earlier one has lower average time value than the later one. In math, $\exists i : |B_{i,m_i}| < |B_{i+1,1}|$. Then, swapping $|B_{i,m_i}|$ and $|B_{i+1,1}|$ in exit sequence increases the delay welfare loss of the former group by $n_{i+1,1} \sum_{j \in B_{i,m_i}} v_j = n_{i+1,1} n_{i,m_i} |B_{i,m_i}|$ while the latter group gains $n_{i+1,1} n_{i,m_i} |B_{i+1,1}|$. Because the former is smaller than the latter, while the delay of all other vehicles is unchanged, we

have a welfare gain which compromises optimality of the initial exit sequence. Therefore, the averages of all sets in the optimal exit sequence are weakly decreasing:

$$|B_{1,1}| \geq \cdots \geq |B_{1,m_1}| \geq |B_{2,1}| \geq \cdots \geq |B_{2,m_2}| \geq \cdots \geq |B_{k,1}| \geq \cdots \geq |B_{k,m_k}|. \qquad (3)$$

Is the first set of Left vehicles from the optimal exit sequence, $B_{1,1}$, identical to the first set in the average-down partitioning of Left-road vehicles, G_1^L? Because they both include the first Left vehicle L_1, one of them should include all elements of the other. Suppose first that $G_1^L \subset B_{1,1}$. By Lemma 1 applied to $B_{1,1}$, we have that $|G_1^L| < |B_{1,1}|$, which contradicts the definition of G_1^L. Second, suppose that $B_{1,1} \subset G_1^L$. By Lemma 1, we then have that $|B_{1,1}| < |G_1^L \backslash B_{1,1}|$. At the same time, $G_1^L \backslash B_{1,1}$ consists of sets or subsets $B_{1,j}$, each of which has an average value no greater than $|B_{1,1}|$, which is a contradiction. Thus, we can conclude $B_{1,1} = G_1^L$, i.e. vehicles in G_1^L exit before all other vehicles.

By iterating the same logic for vehicles remaining after exit of G_1^L, we complete the proof.
∎

3.1.2. Properties of optimal exit sequence

How does the optimal exit sequence change when one optimally chosen vehicle exits the system? Because such exit does not provide any new information, the exit sequence should remain unchanged. The following Proposition supports this hypothesis. It is focused on the Left road without loss of generality.

Proposition 1. *In the average-down partitioning of all vehicles on Left road, exit of the first vehicle L_1 may split the remainder of the first set G_1^L into several sets $G_{1.1}^L, \ldots, G_{1.m}^L$, with average values exceeding the original one:*

$$|G_{1.1}^L| \geq \cdots \geq |G_{1.m}^L| > |G_1^L|.$$

Further sets $G_2^L, \ldots, G_k^L$ remain unchanged.

If the exit sequence is optimal, the fact that L_1, rather than R_1, exits means that $|G_1^L| \geq |G_1^R|$.

Proposition 1 states that, after L_1 exits, the priority of remaining elements of G_1^L over G_1^R is not only preserved but also reinforced, while the priority of other sets is unchanged.

Proof. The fact that G_1^L may split is illustrated by the following example. Suppose G_1^L consists of three elements with values $\{v_1^L, v_2^L, v_3^L\} = \{0, 7, 5\}$. It is a set because $0 < \frac{1}{2}(0+7) < \frac{1}{3}(0+7+5)$. Exit of the first vehicle splits the remainder of G_1^L into two sets, 7 and 5, because $7 > \frac{1}{2}(7+5)$.

The average values of newly created sets are non-increasing by construction of average-down partitioning. The last such set $G_{1.m}^L$ consists of the last elements of G_1^L, thus $|G_{1.m}^L| > |G_1^L|$ by Lemma 1. Appending vehicles from $G_2^L, G_3^L, \ldots$ to $G_{1.m}^L$ cannot increase the average value of the latter, thus impossible in the average-down algorithm. Therefore, newly created sets are made of remaining elements of G_1^L only, and further sets $G_2^L, \ldots, G_k^L$ are unchanged. ∎

How does the average-down partitioning $G_1^L, \ldots, G_k^L$ change when a new vehicle with value $v_{n_L+1}^L$ is added at the end of the queue? The following *updating algorithm* can be proposed.

Initiation Define the *last index, $m = k+1$*. Define the *last set G_m^L* that consists only of the new vehicle L_{n_L+1}.

Updating If the last set is also the first, $m = 1$, stop updating. Otherwise, compare $|G_{m-1}^L|$ and $|G_m^L|$. If the former is greater or equal, stop updating. Otherwise, append G_m^L to G_{m-1}^L so the latter becomes the last set, decrease the last index m by one, repeat the updating cycle.

It is straightforward to show that the outcome of the updating algorithm is the new average-down partitioning.

Proposition 2. *The arrival of a new vehicle to the end of a queue, L_{n_L+1}, cannot delay the exit time of other vehicles on the same road, $L_1, \ldots, L_{n_L}$.*

Proof. By Theorem 1, the exit time of vehicle L_j, belonging to some set G_i^L in the average-down partitioning, depends on the comparison of average value $|G_i^L|$ to average values of groups on the Right road, $|G_1^R|, \ldots, |G_{k_R}^R|$, with higher-value groups exiting first. According to the updating algorithm above, the arrival of L_{n_L+1} may increase, but not decrease, the average value of the set to which L_j belongs. The arrival of L_{n_L+1} has no impact on average values of groups on the Right road. Therefore, the arrival of L_{n_L+1} may expedite, but not delay, any vehicle $L_j, j \leq n_L$. ∎

3.2. Queues with gaps

Vehicles may arrive with gaps, i.e. intervals exceeding the minimum safe distance between them. Moreover, in case of two merging roads and random arrivals over a long period of time, such gaps are necessary to limit from above the expected queue lengths. Queueing of vehicles near the intersection causes these gaps to close, so all vehicles arriving to the intersection can be divided into two groups:

Definition 1. *A vehicle is* proximate *if that vehicle, as well as all vehicles ahead on the same road, have no gaps ahead of them. A vehicle is* distant *if it is not proximate.*

Whenever there is no exit from road S during some time interval of length t, proximate vehicles on that road must be delayed for t units of time, while distant vehicles may keep moving for some time and be delayed, if at all, for a shorter time interval.

3.2.1. Examples

Distant vehicles cannot be ignored when calculating the optimal exit sequence. Consider the following

Example 1. *On the Left road, there is a proximate vehicle L_1 with time value $v_1^L = 1$, followed by a gap of length 0.8, followed by a distant vehicle L_2 with $v_2^L = 8$. On the right, there is a bunch of two proximate vehicles $\{R_1, R_2\}$ with time values $\{0, 4\}$.*

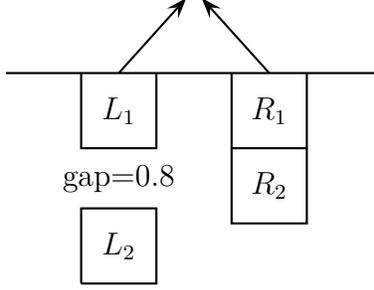The example is illustrated on Figure 1. In Example 1, if the distant vehicle L_2 is ignored,

Figure 1: Layout of vehicles in Example 1. Arrows indicate direction of travel. Horizontal line is intersection threshold.

optimal exit sequence is described by Section 3.1: the two Right vehicles form a set G_1^R with average time value 2, thus they should be given priority over L_1. However, by the time R_1 exits, the gap ahead of L_2 closes, the two Left vehicles form a set G_1^L with average value of 4.5, thus they should be given priority over the remaining Right vehicle R_2. Eventually, the exit sequence is R_1, L_1, L_2, R_2. It is immediate to verify that swapping R_1 and L_1 decreases delay welfare losses, hence ignoring distant vehicles may lead to a suboptimal exit sequence.

At the same time, when accounting for distant vehicles, the algorithm of Section 3.1 cannot be directly applied. This is because the key property used to prove Theorem 1, that a vehicle delay depends only on its rank in the exit sequence, is no longer valid. Consider again Example 1; the optimal exit sequence can be shown to be L_1, R_1, L_2, R_2. Swapping R_1 and L_2 in the exit sequence does not change the rank of R_2 in the exit sequence, but it does increase the time delay of R_2 by 0.8, because both R_1 and R_2 are now delayed not only while L_1 and L_2 are exiting, but also while L_2 is approaching the intersection.

Moreover, not only the algorithm of Section 3.1 is invalidated by the presence of the distant vehicles, but so is its most basic property, Proposition 2, that newly arriving vehicles cannot delay earlier vehicles on the same road. Consider the following

Example 2. *On the Left road, there is a gap of length 0.3, followed by a bunch of distant vehicles L_1, L_2 with values $\{2, 1\}$. On the Right road, there is a proximate vehicle R_1 with value $v_1^R = 1$.*

This layout is illustrated on Figure 2. Suppose first there was no L_2. If L_1 is given priority,
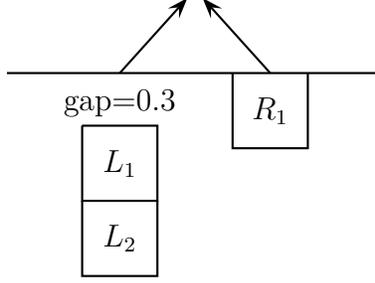
12

Figure 2: Layout of vehicles in Example 2. Arrows indicate direction of travel. Horizontal line is intersection threshold.

R_1 is delayed by 1.3 units of time with equal welfare loss. If R_1 is prioritized, L_1 is delayed by 0.7 units of time with welfare loss of 1.4. Thus, priority of L_1 is optimal. Consider now the full layout with L_2. If L_1 still exits first, R_1 still loses 1.3 units of value; then there is a tie between L_2 and R_1 which results in a further welfare loss of 1, regardless of who exits first. Thus, the total welfare loss is 2.3. At the same time, if R_1 exits first, the bunch $\{L_1, L_2\}$ is delayed by 0.7 units of time, resulting in welfare loss of $0.7 \times (2 + 1) = 2.1$. The second option is better, so we conclude that the presence of vehicle L_2 behind L_1 has delayed the latter. To complicate matters further, L_1 would *not* be delayed by arrival of L_2 if the time value of the latter was much lower (e.g. $v_2^L = 0$) or much higher (e.g. $v_2^L = 10$).

Given these examples, there seems to be no way to find the optimal exit sequence other than to build a decision tree and to compare multiple exit sequences between each other. The algorithm must be fast, as vehicles arrive and await exit commands in real time. The results of Section 3.1, as well as some additional results below, help to refine the set of sequences to be compared. In many cases, it is possible to determine which of the two vehicles $\{L_1, R_1\}$ receives priority without combing through the entire decision tree.

3.2.2. Definitions and properties

To facilitate further analysis, we now clarify some definitions that are used to calculate congestion welfare losses.

Definition 2. Exit gap *is a gap ahead of a vehicle that does not close before that vehicle*

clears the intersection.

No vehicle exits the intersection during an exit gap.

The above examples of this section implicitly made a natural assumption that delay of a vehicle is the difference between the true exit time and the exit time if that vehicle exited without slowing down, which we refer to as the "simple" definition. Below, we make use of a less intuitive but more mathematically convenient "formal" definition:

Definition 3. Delay *of a vehicle is the difference between (i) the true exit time and (ii) the exit time if the vehicle* was proximate *(i.e. if all gaps on the same road ahead did not exist) and exited without slowing down.*

Furthermore, delay of a vehicle can be divided into two types.

Definition 4. Rank delay *of a vehicle occurs during the time when a vehicle from the opposite road is exiting.* Gap delay *of a vehicle occurs during the time when there is an exit gap.*

The total delay of a vehicle is the sum of its rank and gap delays. Whenever a Right vehicle is exiting the intersection, the rank delay of all remaining Left vehicles is incremented. Whenever there is an exit gap, the gap delay of all remaining vehicles, both Left and Right, is incremented.

The simple and the formal definitions of delay result in identical comparisons of exit sequences. Consider Example 1, in which the optimal exit sequence is L_1, R_1, L_2, R_2. According to the simple definition, the delays are[4] $\{0, 1, 0.2, 2\}$ with the associated welfare loss of $\{0, 1, 0.2, 2\}^T \times \{1, 0, 8, 4\} = 9.6$, where the latter vector is that of vehicle time values. Under the formal definition, there are no gap delays (because there are no exit gaps), while rank delays are $\{0, 1, 1, 2\}$ and the associated welfare loss is 16. Consider the alternative exit sequence: L_1, L_2, R_1, R_2, which implies an exit gap of length 0.8 ahead of L_2. The simple-definition delays are $\{0, 0, 2.8, 2.8\}$, with welfare loss of $\{0, 0, 2.8, 2.8\}^T \times \{1, 8, 0, 4\} =$

---

[4]All vectors in this example are sorted in order of vehicle exit.

11.2. Under the formal definition, the rank delays are $\{0, 0, 2, 2\}$ while the gap delays are $\{0, 0.8, 0.8, 0.8\}$. The welfare loss due to all delays is then 17.6. While the formal definition results in different values of the welfare loss, the comparison between two exit sequences is the same $(17.6 - 16 = 11.2 - 9.6)$.

The following Proposition is a weaker version of Theorem 1 that facilitates calculation of optimal exit sequence.

Proposition 3. *Apply the average-down partitioning to the Right vehicles, which results in sets $\{G_1^R, \ldots, G_{k_R}^R\}$. Take an arbitrary $n \leq n_L$ and apply the average-down partitioning to Left vehicles $L_1, \ldots, L_n$, which results in sets $\{G_1^L, \ldots, G_{k_L}^L\}$. If members of G_1^L are all proximate and $|G_1^L| \geq |G_1^R|$, then the last member of G_1^L, denoted L_g, exits before R_1.*

Proposition 3 is especially useful at the time of high congestion, when there are many proximate vehicles on both roads and thus the condition of the Proposition, that all members of G_1^L are proximate, is satisfied in a greater proportion of cases.

Proof. Suppose the contrary, that vehicles $R_1, \ldots, R_m$ exit before L_g. Denote by $h_i \geq 0$ the length of an exit gap immediately ahead of R_i, and by $h^a = \sum_{i=1}^m h_i$ the total length of exit gaps ahead of L_g. Putting all these Right vehicles behind L_g would have the following welfare effects:

- Rank delays of Left vehicles are reduced while rank delays of Right vehicles are increased by equal amount. Because $|G_1^L| \geq |G_1^R| \geq \frac{1}{m} \sum_i R_i$ (where the second inequality is by definition of G_1^R), the overall welfare due to changes in rank delays weakly increases. The proof is analogous to that in Theorem 1.

- Exit delays for all vehicles in G_1^L are now eliminated. Because each of vehicles $R_1, \ldots, R_m$ now has to yield to each member of G_1^L, some gaps ahead of these vehicles close before they exit, so the sum of the exit gaps ahead of Right vehicles decreases down to some $h^b \in [0, h^a)$. Because the exit time of the entire group $G_1^L, R_1, \ldots, R_m$ is reduced, the total delay of each vehicle behind R_m cannot increase. Thus, we conclude that the

15

welfare effect of changing gap delays is positive (strictly if $h^a > 0$), because such gaps delays are reduced to zero for vehicles in G_1^L, are strictly reduced for $R_1, \ldots, R_m$, and are weakly reduced for all subsequent vehicles.

Thus, the reshuffling is welfare-improving, contradicting optimality of $R_1, \ldots, R_m$ being ahead of L_g. ∎

3.2.3. The exit sequence algorithm

This section builds a decision tree that helps to find the optimal exit sequence in the presence of gaps ahead of vehicles. Vehicle layouts are equivalent to nodes in the decision tree. A vehicle exit sequence is then an equivalent to a decision tree path. The objective is to find a sequence with a minimal aggregate welfare loss.

We now specify an algorithm for finding such sequence. The initial layout is the position of all known vehicles at the time the decision is being made. At each layout, the following sequence of decisions is made.

0. If there are gaps ahead of both L_1 and R_1, fast-forward until at least one vehicle becomes proximate, proceed to the next layout.

1. If the gap ahead of R_1 (L_1) is greater or equal to one, L_1 (R_1) exits first as such exit does not delay any of the Right (Left) vehicles. Proceed to the next layout.

2. Apply the average-down partitioning to both Left and Right vehicles, ignoring gaps ahead of them. In all subsequent steps, we will assume $|G_1^L| \geq |G_1^R|$ without loss of generality. If all vehicles in G_1^L are proximate, then G_1^L receives priority by Proposition 3 and exits; proceed to the new layout.

3. If some but not all vehicles in G_1^L are proximate, apply the average-down partitioning to those proximate vehicles, resulting in some $\{G_{1.1}^L, \ldots, G_{1.m}^L\} \subset G_1^L$. If $|G_{1.1}^L| \geq |G_1^R|$, then $G_{1.1}^L$ receives priority by Proposition 3 and exits; proceed to the new layout.

16

4. If some vehicles in G_1^L are proximate but $|G_{1.1}^L| < |G_1^R|$, consider two possible paths:

   (a) R_1 exits first, reducing gaps on the Left road;

   (b) Left vehicle(s) exit first. The exact number of exiting vehicles is determined as follows. If there are gaps ahead of any Right vehicles, only L_1 exits, as such exit, by closing the Right gaps, may entail priority of the Right vehicles. Otherwise, proximate Left vehicles can exit only in groups $G_{1.1}^L, \ldots, G_{1.m}^L$. Compare $|G_{1.1}^L|$ and v_1^R. If $|G_{1.1}^L| \geq v_1^R$, $G_{1.1}^L$ is the group to exit, as it may be better than R_1 exiting first. If $|G_{1.1}^L| < v_1^R$, proximate Left vehicles alone cannot be prioritized over R_1, only in combination with distant Left vehicles. In such case, all proximate $\{G_{1.1}^L, \ldots, G_{1.m}^L\}$ and the first distant Left vehicles exit, with an exit gap in between.

5. If no vehicles in G_1^L are proximate, two possible paths should be considered:

   (a) R_1 exits first, reducing gaps on the Left road;

   (b) L_1 exits first with an exit gap ahead of it.

Some sequences may converge to the same intermediate layout, allowing to drop sequences with higher welfare losses. For example, if both L_1 and R_1 have no gaps ahead of them, then two exit sequences, $\{L_1, R_1\}$ and $\{R_1, L_1\}$, result in the same layout after two periods of time. If $v_1^L > v_1^R$, the latter sequence can be dropped as it entails higher welfare losses. Dropping sequences may result in a situation when all remaining exit sequences under consideration prioritize the same vehicle from the set $\{L_1, R_1\}$, meaning that this prioritized vehicle can be allowed to exit before the entire decision tree has been analyzed.

To summarize, the proposed exit sequence algorithm builds a decision tree with vehicle layouts as nodes. The tree splits into two branches only at some nodes that feature gaps ahead of some vehicles. In many nodes, Proposition 3 leaves only one possible branch. In addition, by allowing Left (Right) vehicles to exit in sets, Proposition 3 may close many

Right (Left) gaps, further reducing the problem dimensionality. Also, the tree branches may converge, allowing to discard some tree paths before reaching the end of the tree. All these shortcuts enable to drop the vast majority of $\frac{(n_L+n_R)!}{n_L!n_R!}$ possible tree branches.

3.3. Limited planning horizon

Despite the best effort to speed up priority decision-making, the authority may have to announce a priority decision before the optimal exit sequence has been found. A natural solution is to limit the planning horizon (in terms of distance to the intersection) of the authority and ignore vehicles beyond that horizon. The goal of this section is to analyze how much expected welfare can be lost due to such limit.

For transparency of our analysis, we limit ourselves to the case of queues without gaps and without any expected additional arrivals. The time values of vehicle occupants in the queues are assumed to be i.i.d.[5] In numerical examples below, we assume the exponential distribution with unitary mean. With n_S vehicles queued on road $S \in \{L, R\}$, the total amount of delay is $n_L n_R$ vehicle-time units, thus the average delay per vehicle is $\frac{n_L n_R}{n_L+n_R}$.

Consider first zero planning horizon, i.e. the authority completely ignoring time value signals and randomizing the exit priority. Throughout the rest of the paper, we will refer to this as *benchmark mechanism,* with which mechanisms of this paper will be compared. In the benchmark mechanism, the expected delay of a vehicle does not depend on the occupants' time value, so the expected welfare lost to congestion is $n_L n_R E v$.

Next, consider the layout with one vehicle on each road, $n_L = n_R = 1$. Then, the mechanism of allocating priority proposed in this paper is essentially an all-pay auction with two bidders. The optimal exit sequence is to merely give priority to the higher-value vehicle,

---

[5]While Section 2 explicitly assumes i.i.d. values of those entering the system, the values of those present in the queue at a given time may be interdependent due to non-random nature of previous exit from the queue. In this section, we essentially assume that there was no previous exit.

and the expected delay welfare loss is

$$\int_0^\infty \left[\int_0^{v_1^L} v_1^R \mathrm{d}F(v_1^R) + v_1^L(1 - F(v_1^L)) \right] \mathrm{d}F(v_1^L),$$

where $F(\cdot)$ is the c.d.f. of time values. Assuming $F(x) = 1 - \exp(-x)$, the expected delay welfare loss is $\frac{1}{2}$, compared to the loss of 1 in the benchmark of ignoring vehicle values.

The next example assumes $n_L = 2$ and $n_R = 1$, which is the largest number of vehicles allowing to illustrate exit sequences on a two-dimensional graph. What happens if the planning horizon is unity, i.e. if the value of the second Left vehicle is initially ignored? Then, the incorrect exit sequence may result only in one case: R_1 exits first $v_1^R > v_1^L$ while it should exit last $v_1^R < \frac{1}{2}(v_1^L + v_2^L)$. Figure 3 illustrates the value realizations resulting in the incorrect exit sequence. Assuming exponential value distribution with unitary mean, the
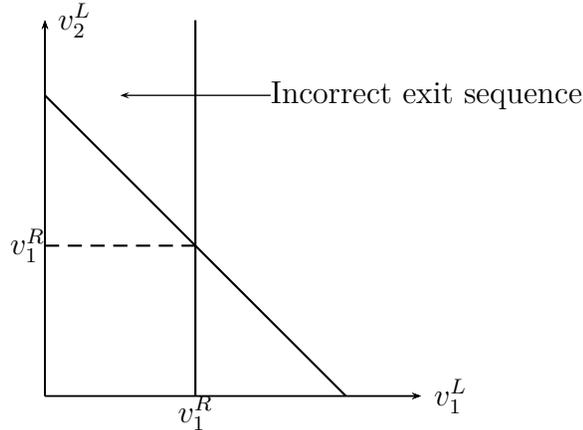

Figure 3: Incorrect exit sequence due to limited planning horizon

expected delay welfare loss under the benchmark is 2. Under unitary planning horizon, it is

$$\int_{v_1^R=0}^\infty \left\{ \int_{v_1^L=0}^{v_1^R} (v_1^L + Ev_2^L)\mathrm{d}F(v_1^L) + (1 - F(v_1^R)) \left[v_1^R + \int_{v_2^L=0}^{v_1^R} v_2^L \mathrm{d}F(v_2^L) + v_1^R(1 - F(v_1^R)) \right] \right\} \mathrm{d}F(v_1^R),$$

equal to $\frac{7}{6}$. Under the full planning horizon of two vehicles, the expected welfare loss is

reduced down to

$$\int_{v_1^L=0}^{\infty} \int_{v_2^L=0}^{v_1^L} \left[\int_{v_1^R=0}^{v_2^L} 2v_1^R \mathrm{d}F(v_1^R) + \int_{v_1^R=v_2^L}^{v_1^L} (v_1^R + v_2^L)\mathrm{d}F(v_1^R) + (v_1^L + v_2^L)(1 - F(v_1^L)) \right] \mathrm{d}F(v_2^L)\mathrm{d}F(v_1^L)$$
$$+ \int_{v_1^L=0}^{\infty} \int_{v_2^L=v_1^L}^{\infty} \left[\int_{v_1^R=0}^{\frac{1}{2}(v_1^L+v_2^L)} 2v_1^R \mathrm{d}F(v_1^R) + (v_1^L + v_2^L)(1 - F(\frac{1}{2}(v_1^L + v_2^L))) \right] \mathrm{d}F(v_2^L)\mathrm{d}F(v_1^L),$$

equal to $\frac{19}{18}$.

As evidenced from the above formulas, the complexity of analytical computation rapidly increases with the number of queued vehicles. For larger number of vehicles, we replace the precise analytical results by approximations calculated from simulations of vehicle exit. We consider two examples. In the first example, initially there are $N = 20$ vehicles queued on each entry road. In the second example, N is increased to 50. In each of the two examples, 10000 initial vehicle layouts were simulated. We continue to assume i.i.d. exponential distribution of values. In each simulated layout, vehicle exit sequences under planning horizons from 1 to N were calculated, and the associated delay welfare losses recorded. Figure 4 reports the results averaged across the simulations, relative to those of the benchmark mechanism. Define by *frontier* the welfare loss under the maximum possible planning horizon N, i.e. when all queued vehicles are accounted for from the beginning. A planning horizon of only ten vehicles reduces the gap between the benchmark and the frontier by 92% when $N = 20$, and by 75% when $N = 50$. Thus, we can conclude that even relatively short planning horizons can bring welfare close to the frontier.

Also note that introducing gaps between vehicles would make it less likely that vehicles behind the gaps (i.e. distant) are pivotal in the optimal exit sequence, making the welfare of a limited planning horizon even closer to the frontier.
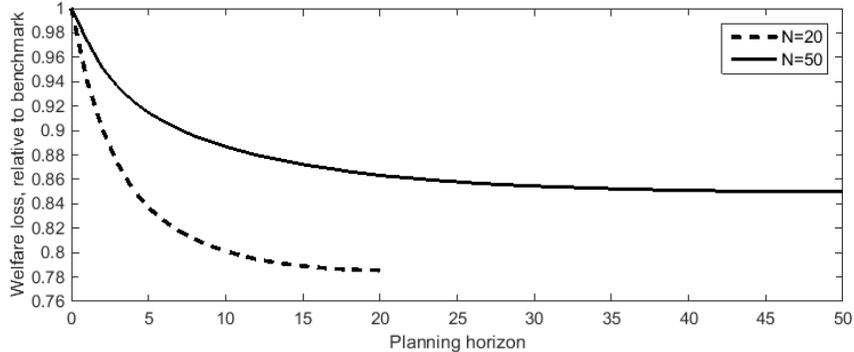
Figure 4: Planning horizon and expected delay welfare loss

4. Optimal tolls

The value of time of vehicle occupants is their private information. The authority can make optimal exit sequence decisions only if it has correct estimates of the time values. The time values can be elicited by an optimally constructed toll, using the principles of mechanism design.

We will assume that each traveler, prior to arriving to the intersection, reports her type $\tilde{v}$ and pays a toll which is a function of the reported type, $p(\tilde{v})$. We assume that the toll does not depend (i) on whether the traveler has arrived from Left or from Right, and (ii) on the reported types of other vehicles approaching the intersection.

The assumption (i) is motivated by the idea that a typical trip requires crossing multiple intersections, and the value report is made only once, for all intersections, for convenience of the travelers. In our stylized model with a single intersection, we assume that the direction of arrival, Left or Right, is random for each given vehicle, with identical and independent probability distribution for all vehicles.

The assumption (ii) is motivated by the idea that the value report is made before the identity of other travelers at a given intersection is known.

Denote by $t(v_i)$ the expected delay (including the rank and gap delays, as in Definition 4) of vehicle i, assuming the authority prescribes the optimal exit sequence for vehicles within

21

a certain planning horizon. We have the following result.

Proposition 4. *The expected delay $t(\cdot)$ is weakly decreasing in time value. In math, for any v_i, v_i^a such that $v_i < v_i^a$, we have $t(v_i) \geq t(v_i^a)$.*

Proof. Consider a layout of vehicles l in which the vehicle in question i has just crossed the authority planning horizon. There are N^l vehicles within the planning horizon on both roads, including i, with values $\mathbf{v}^l = \{v_1^l, \ldots, v_N^l\}^T$. There are K^l exit sequences available, such that the vector of total (i.e. both rank and gap) delays for all vehicles in exit sequence k is[6] $\mathbf{t}_k^l = \{t_{k,1}^l, \ldots, t_{k,N^l}^l\}^T$. Suppose the optimal exit sequence is $k(l)$ so that

$$(\mathbf{t}_{k(l)}^l)^T \mathbf{v}^l \leq (\mathbf{t}_k^l)^T \mathbf{v}^l, \forall k = 1, \ldots, K^l. \tag{4}$$

Suppose now the time value of vehicle i is increased from v_i to v_i^a. The total welfare loss in exit sequence k is increased by $t_{k,i}^l(v_i^a - v_i)$. For exit sequences k in which i exits later than previously optimal, such that $t_{k(l),i}^l \leq t_{k,i}^l$, inequality (4) is preserved so they cannot become better than $k(l)$. For exit sequences k in which i exits sooner, $t_{k(l),i}^l > t_{k,i}^l$, inequality (4) may be reversed, so such exit sequences may become optimal. Thus, we conclude that increased value of time v_i weakly decreases the delay of i in every given vehicle layout.

The expected delay $t(v_i)$ is the expectation of vehicle i delay over all possible layouts, $t(v_i) = E_l t_{k(l),i}^l$. Such expectation weakly decreases in v_i because (i) so does every element of the expectation, and (ii) the expectation is calculated over layouts l in which the vehicle i has just arrived and therefore v_i had no effect on the probability distribution of other vehicles in the layouts. ∎

We now calculate the toll function $p(\cdot)$ that induces truthful revelation of time value v. This is a straightforward application of mechanism design principles, and is similar to

---

[6]Because the exit sequence may be updated as new vehicles appear within the planning horizon, it might be more appropriate to calculate the *expected* delay. We abstract from this problem for the sake of exposition clarity, assuming there will be no new arrivals. The Proposition remains valid if new arrivals are accounted for.

Theorem 1 of Myerson and Satterthwaite (1983). For a traveler with true value of time v and reported value of time $\tilde{v}$, the welfare loss due to both delays and toll payment is $w(\tilde{v}, v) \equiv t(\tilde{v})v + p(\tilde{v})$. For truthful telling, we should have

$$w(v_1, v_1) = t(v_1)v_1 + p(v_1) \leq t(v_2)v_1 + p(v_2) = w(v_2, v_1),$$

$$w(v_2, v_2) = t(v_2)v_2 + p(v_2) \leq t(v_1)v_2 + p(v_1) = w(v_1, v_2),$$

for any $v_1, v_2 \geq 0$. By manipulating with these, we obtain

$$t(v_1)(v_1 - v_2) \leq w(v_1, v_1) - w(v_2, v_2) \leq t(v_2)(v_1 - v_2),$$

which implies

$$w(v, v) = p_1 + \int_{z=0}^{v} t(z)\mathrm{d}z. \tag{5}$$

By comparing (5) against the definition of $w(\tilde{v}, v)$ above, we obtain

$$p(v) = p_1 - t(v)v + \int_{z=0}^{v} t(z)\mathrm{d}z. \tag{6}$$

Wherever $t(\cdot)$ discontinuously decreases, $p(\cdot)$ discontinuously increases. Wherever $t(\cdot)$ is continuous, we have $p'(v) = -t'(v)v \geq 0$. The constant p_1 is the *baseline toll* by a traveler with zero time value. If vehicle arrivals are exogenous, increasing p_1 simply redistributes welfare from travelers to the traffic authority. If vehicle arrivals depend on expected welfare generated by such travel, increasing p_1 reduces arrivals, which also affects the function $t(\cdot)$.

By a *traffic mechanism* we will denote the combination of the optimal exit sequence and the toll function given by (6).

5. Expected welfare implications of traffic mechanisms

This section studies various welfare aspects of optimal exit regulation in the presence of asymmetric information. Because such regulation requires travelers to pay tolls for exit priority, we need to specify how toll revenues are spent by the authority. We follow the convention that all revenues are rebated equally to all travelers, either in cash or in terms of public good provision.

While application of traffic mechanisms undoubtedly increases aggregate welfare, some groups of travelers might be hurt. Redistribution from winners to losers is impossible due to the private nature of time valuations. We now analyze who might be hurt by such regulation.

5.1. Exogenous arrivals

Suppose the distribution of time values in the population is given by p.d.f. $f(\cdot)$, and the entire population travels through the intersection regardless of costs. Under a traffic mechanism, the rebate of the authority to travelers is $Ep \equiv \int_{v=0}^{\infty} p(v)f(v)\mathrm{d}v$, and the expected welfare loss of a traveler with value v is (cf.(5)) $w_1(v) = t(v)v + p(v) - Ep$.

We now compare the traffic mechanism to the benchmark mechanism which ignores traveler signals. Assuming the traffic mechanism delivers exit sequences without exit gaps, the average exit times in both traffic and benchmark mechanisms are the same, thus the latter offers the expected exit time of $Et = \int_{v=0}^{\infty} t(v)f(v)\mathrm{d}v$ to all travelers regardless of their time value. The associated welfare loss is $w_0(v) = vEt$.

How do $w_1(v)$ and $w_0(v)$ compare to each other?

Proposition 5. *Assuming exogenous arrivals and no exit gaps in the traffic mechanism, such mechanism is Pareto-improving relative to the benchmark.*

Figure 5 illustrates the comparison of the delay welfare losses under the two mechanisms.

Proof. For both mechanisms being compared, calculate the derivatives of welfare loss with respect to time value: $w_1'(v) = t(v)$ and $w_0' = Et$. The former decreases with v while the
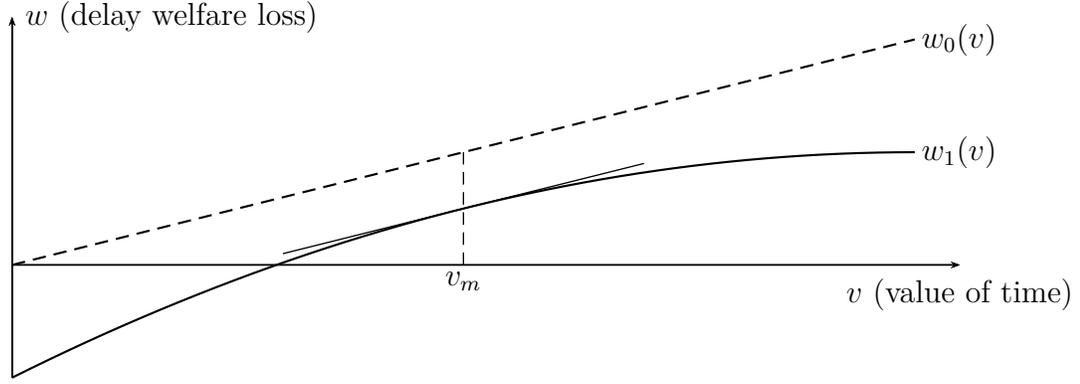
Figure 5: Expected welfare losses, as function of time value, under benchmark (w_0) and traffic (w_1) mechanisms with exogenous arrivals

latter is constant, therefore the maximum of the difference $w_1(v) - w_0(v)$ is achieved for traveler with value v_m such that $t(v_m) = Et$. To complete the proof, it suffices to show that

$$w_1(v_m) = t(v_m)v_m + p(v_m) - Ep \overset{?}{\leq} v_m Et = w_0(v_m),$$

which is equivalent to showing that

$$p(v_m) \overset{?}{\leq} Ep. \tag{7}$$

Because v is a realization of a random time-value variable, so is $y = t(v)$. The expectation of the latter is $Ey = t(v_m)$. Denote by $s(\cdot)$ the inverse function of $t(\cdot)$, such that $s(t(v)) \equiv v$. Because $t(\cdot)$ is weakly decreasing and (possibly) discontinuous, so is $s(\cdot)$. Denote $q(y) \equiv p(s(y)) = p_1 + \int_{z=y}^{t(0)} s(z)\mathrm{d}z$. We have that $q'(y) = -s(y)$ is weakly increasing, thus $q(y)$ is a convex function, thus

$$p(v_m) = q(Ey) \leq Eq(y) = Ep$$

by Jensen's inequality, which proves (7). ∎

Therefore, if the traffic mechanism does not alter the number of travelers, relative to the benchmark, and does not generate exit gaps, it will benefit each and every traveler on the

road.

5.2. Endogenous arrivals and traffic volume

We now assume that travel through the intersection is a matter of choice with non-trivial outside opportunities. This section provides a sketch of how welfare and travel decisions change following the implementation of a traffic mechanism.

For that purpose, suppose the travelers differ not only in their value of time v but also in the value y of traveling through the intersection. Then, a traveler who was delayed for t units of time and paid p for getting through enjoys welfare $u = y - tv - p$, while the welfare when not traveling (or using another method of transportation) is normalized to zero. Assuming that under mechanism i the expected delay and toll payment are equal to $t_i(v)$ and $p_i(v)$, respectively, travelers enter the intersection under such mechanism iff their value y of getting through exceeds $y_i(v) = t_i(v)v + p_i(v)$. The roll revenues are rebated equally to all population, regardless of whether they enter the intersection or not.

Under the benchmark mechanism, the authority is assumed to charge a uniform toll p_0 which, by regulating entry and the expected delay t_0, maximizes aggregate welfare. The entry cutoff is then a linear function of time value, $y_0(v) = t_0v + p_0$.

Section 5.1 demonstrates that, if under the traffic mechanism, entry decisions and the total toll revenues were the same as in the benchmark, then all those who travel would be strictly better off. But that would have the following effects on the equilibrium. Strictly higher welfare from travel would lower the cutoff $y_1(\cdot)$ and thus induce more travelers to enter, causing more traffic. More traffic would increase both delays $t_1(\cdot)$ and tolls p_1, offsetting part of the initial welfare gain.

For a minority of travelers, those with time values close to v_m as defined in Section 5.1, the negative effects of higher traffic may exceed the positive effects of the optimal exit sequence, making them worse off from the traffic mechanism. In other words, extra traffic

of those with low or high value of time may crowd out those in the middle.

At the same time, low-y travelers who do not travel under both mechanisms are better off due to increased toll revenues that are rebated to them.

Figure 6 illustrates the effects of a traffic mechanism implementation in the environment with endogenous entry.
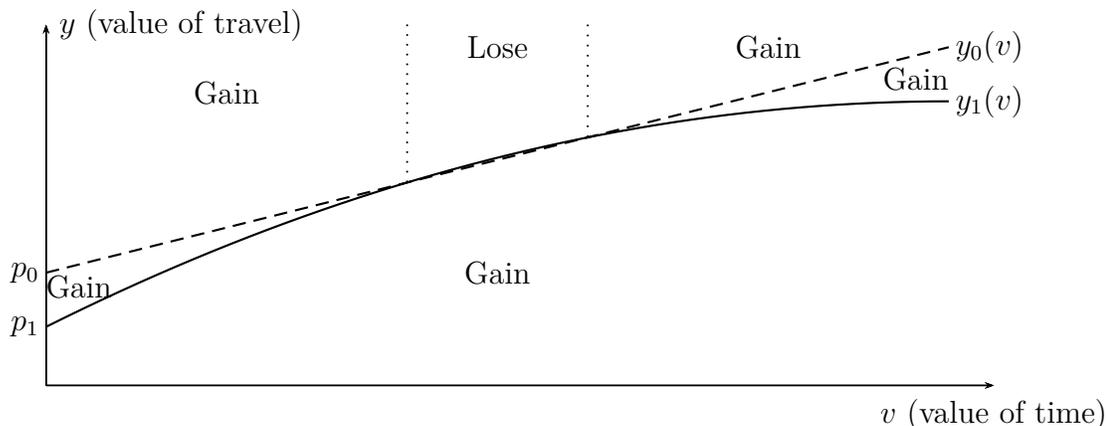

Figure 6: Entry cutoff, as function of time value, under benchmark (y_0) and traffic (y_1) mechanisms with endogenous arrivals

5.3. First-best optimality of traffic mechanism

It is well-known that asymmetric information can lead to inefficient outcomes. For that reason, the mechanism design literature operates with the concept of the second best, i.e. maximum aggregate welfare that can be achieved accounting for asymmetry of information. This section shows that the second best traffic mechanism is also the first best. In other words, no welfare is lost due to private nature of time values.

With exogenous arrivals, this conclusion is quite trivial: the mechanism induces travelers to truthfully reveal their time values, and implements the first-best exit sequence given these values. There is no channel through which the aggregate welfare can be lost.

In case of endogenous arrivals, however, the mechanism also affects entry decisions which are not necessarily optimal. The application of the mechanism can potentially result in too

much or too little traffic, or in a different distribution of traffic by type, compared to the first-best. The following lemma is instrumental in proving the opposite.

Lemma 2. *In the traffic mechanism, the toll function $p(\cdot)$ equals, up to a constant, to the expected externality that that vehicle has on other vehicles.*

Proof. Consider a specific traffic layout with n vehicles. Partition the space of vehicle time values $\mathbf{v} = \{v_1, \ldots, v_n\}^T$ into $\{S_1, \ldots, S_K\}$ such that an exit sequence k is optimal iff $\mathbf{v} \in S_k$. Denote the exit times in exit sequence k by $\mathbf{t}^k = \{t_1^k, \ldots, t_n^k\}^T$. If a vector of values $\mathbf{v}$ is a boundary of some two subsets S_k, S_l, then

$$\mathbf{v}^T \mathbf{t^k} = \mathbf{v}^T \mathbf{t^l}. \tag{8}$$

For proof of (8), suppose the left-hand side is strictly smaller. This inequality would be preserved when changing values from $\mathbf{v}$ to some sufficiently proximate $\mathbf{v}^a \in S_l$, violating the definition of S_l.

Denote by $R_i(\mathbf{v})$ the externality of vehicle i on other vehicles in the traffic layout in question, given the vector of values. How does $R_i(\mathbf{v})$ change when i's value increases from v_i to some $v_i + \epsilon$? Denote $\mathbf{v}^a(\epsilon) = \{v_i + \epsilon, v_{-i}\}$ the new vector of values. If both $\mathbf{v}$ and $\mathbf{v}^a(\epsilon)$ belong to the same S_k for all sufficiently small ϵ, such value increase does not alter the exit sequence and therefore does not change $R_i(\mathbf{v})$. If $\mathbf{v} \in S_k$ and $\mathbf{v}^a(\epsilon) \in S_l, l \neq k$ for any $\epsilon > 0$, then $\mathbf{v}$ is the boundary of S_k and S_l, and $R_i(\mathbf{v})$ changes by $\sum_{j \neq i} v_j(t_j^l - t_j^k)$, which is equal to

$$- v_i(t_i^l - t_i^k) \tag{9}$$

according to (8). The change in $R_i(\mathbf{v})$ is non-negative because $t_i^l \leq t_i^k$ according to Proposition 4. Therefore, $R_i(\mathbf{v})$ is an increasing step-like function of v_i, with steps occurring where exit sequences are changed and having the height given by (9). To calculate the value of $R_i(\mathbf{v})$ at any given v_i, denote $v_i^{[0]} = 0$ and $v_i^{[1]}, v_i^{[2]}, \ldots, v_i^{[k]}, \ldots$ the values of v_i at which the

exit sequence is changed, for a given $\mathbf{v}_{-i}$. Denote the delay of vehicle i as a function of v_i, for a given v_{-i}, by $t(v_i, \mathbf{v}_{-i})$. It is a decreasing (by Proposition 4) step-like function; denote its value between $v_i^{[k]}$ and $v_i^{[k+1]}$ by $t_i^{[k]}$. Then, the externality function $R_i(\mathbf{v})$ for a given $v_i \in [v_i^{[k]}, v_i^{[k+1]})$ can be presented as (cf. (9))

$$
\begin{aligned}
R_i(\mathbf{v}) &= R_i^0(\mathbf{v}_{-i}) - \sum_{l=1}^{k} v_i^{[l]} \left(t_i^{[l]} - t_i^{[l-1]} \right) \\
&= R_i^0(\mathbf{v}_{-i}) + v_i^{[1]} t_i^{[0]} + \sum_{l=1}^{k-1} \left(v_i^{[l+1]} - v_i^{[l]} \right) t_i^{[l]} + (v_i - v_i^{[k]}) t_i^{[k]} - v_i t_i^{[k]} = R_i^0(\mathbf{v}_{-i}) + \int_{z=0}^{v_i} t(z, \mathbf{v}_{-i}) \mathrm{d}z - v_i t(v_i, \mathbf{v}_{-i}).
\end{aligned}
$$

$$(10)$$

By taking the expectation of (10) over all possible $\mathbf{v}_{-i}$, and over all possible traffic layouts, we obtain the expected externality function as follows:

$$
R(v) = R_0 + \int_{z=0}^{v} t(z) \mathrm{d}z - v t(v),
$$

which is equal to (6) up to a constant. ∎

Theorem 2. *The traffic mechanism results in first-best aggregate welfare iff the baseline payment p_0 in (6) equals the expected externality of zero-time-value vehicle on other traffic.*

The proof of the theorem is straightforward: if payment $p(0) = p_1$ by the zero-time-value vehicle equals its expected externality, then by Lemma 2 so does the payment by vehicles with all other values. As all externalities are fully internalized, all travelers make socially optimal entry decisions. After entry, the traffic mechanism induces travelers to truthfully report their types and implements the socially optimal exit sequence.

6. Discussion and conclusion

This paper is a first step in developing traffic regulation mechanisms that respect heterogeneity of the time value of travelers. The paper characterizes some important theoretical

results, such as the optimal exit sequence, the optimal payment for priority, some welfare comparisons, and first-best optimality of the traffic volume and composition by type.

Some critics of this research have argued that auctioning off traffic priority will discriminate the poor. However, as evidenced by Section 5, the minimum payment after government rebates will actually decrease. Those willing to travel at a minimum cost will therefore benefit from the traffic priority mechanism.

Another virtue of auctioning off traffic priority, not discussed in the main body of the paper, is increased carpool incentives. Nowadays, governments enhance such incentives by two primary regulation methods: (i) special lanes, which increase speed of carpoolers but require massive investments, and (ii) road tolls and parking fees, which increase incentives to share the costs of travel. A priority-selling traffic mechanism creates a third incentive: carpoolers can pool their funds not only to save on tolls but also to pay more for priority, thus travel faster even in the absence of expensive and space-consuming special lanes.

This paper has assumed that each traveler has a time-invariant value of time. Alternative assumptions can be considered in future research. For example, some travelers intend to arrive to their destination at a certain moment of time, and are averse to the risk of being late. Their willingness to pay for priority would then increase as the intended arrival time approaches.

The bright future of green lights for sale is ahead.

References

Bergemann, D., Said, M., 2011. Dynamic auctions. Wiley Encyclopedia of Operations Research and Management Science.

Cowan, R., 1979. The uncontrolled traffic merge. Journal of applied probability 16 (02), 384–392.

Feigenbaum, J., Shenker, S., 2002. Distributed algorithmic mechanism design: Recent results and future directions. September.

Kamal, M. A. S., Imura, J.-i., Hayakawa, T., Ohata, A., Aihara, K., 2015. A vehicle-intersection coordination scheme for smooth flows of traffic without using traffic lights. IEEE Transactions on Intelligent Transportation Systems 16 (3), 1136–1147.

Milakis, D., Van Arem, B., Van Wee, B., 2017. Policy and society related implications of automated driving: A review of literature and directions for future research. Journal of Intelligent Transportation Systems, 1–25.

Milanés, V., Godoy, J., Villagrá, J., Pérez, J., 2011. Automated on-ramp merging system for congested traffic situations. IEEE Transactions on Intelligent Transportation Systems 12 (2), 500–508.

Myerson, R. B., Satterthwaite, M. A., 1983. Efficient mechanisms for bilateral trading. Journal of economic theory 29 (2), 265–281.

Nisan, N., Ronen, A., 1999. Algorithmic mechanism design. In: Proceedings of the thirty-first annual ACM symposium on Theory of computing. ACM, pp. 129–140.

Parkes, D. C., Singh, S. P., 2004. An mdp-based approach to online mechanism design. In: Advances in neural information processing systems. pp. 791–798.

Pollatschek, M. A., Polus, A., Livneh, M., 2002. A decision model for gap acceptance and capacity at intersections. Transportation Research Part B: Methodological 36 (7), 649–663.

Said, M., 2012. Auctions with dynamic populations: Efficiency and revenue maximization. Journal of Economic Theory 147 (6), 2419–2438.

Tanner, J., 1962. A theoretical analysis of delays at an uncontrolled intersection. Biometrika 49 (1/2), 163–170.

Wang, Z., Kulik, L., Ramamohanarao, K., 2007. Proactive traffic merging strategies for sensor-enabled cars. In: Proceedings of the Fourth ACM International Workshop on Vehicular Ad Hoc Networks. VANET '07. ACM, New York, NY, USA, pp. 39–48.

URL `http://doi.acm.org/10.1145/1287748.1287755`