ITS UCDAVIS
INSTITUTE OF TRANSPORTATION STUDIES

Research Report – UCD-ITS-RR-10-41

System-wide Ramp Metering as a Policy Tool to Induce Efficient Travel in a Freeway Corridor

March 2010

Wei Shen
Michael Zhang

Institute of Transportation Studies ◦ University of California, Davis
One Shields Avenue ◦ Davis, California 95616
PHONE (530) 752-6548 ◦ FAX (530) 752-6572
www.its.ucdavis.edu

UCDAVIS
SUSTAINABLE TRANSPORTATION CENTER
of the Institute of Transportation Studies

Final Research Report R03-3

System-wide Ramp Metering as a Policy Tool to Induce Efficient Travel in a Freeway Corridor

March 2010

Wei Shen
Michael Zhang

Department of Civil and Environmental Engineering
University of California, Davis
One Shields Ave., Davis, CA 95616

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

1. INTRODUCTION

In a congested transportation network, a traveler's trip can cause additional delays to other travelers who enter the network later. Due to the presence of this congestion externality, the departure time and route choices of travelers who seek to optimize their own travel experience usually lead to inefficient temporal and spatial traffic distribution in a transportation network in terms of social welfare. This 'welfare gap' between the Wardropian user equilibrium (a product of selfish choices) and system optimum traffic patterns can be, in theory, eliminated by internalizing the congestion externality via *marginal cost pricing*.

Considerable effort has been devoted to the understanding, characterization and more importantly elimination of (or narrowing) this welfare gap in the context of morning commute. The seminal work by *(1)* delineates the first dynamic economic model for congestion pricing, which focuses on a single route with one bottleneck. The model assumes that all the travelers have the same desired arrival time and choose their departure times by balancing the tradeoff between travel cost and schedule delay cost (i.e., the cost of early or later arrival). A time-dependent toll is derived to drive the user equilibrium pattern toward a system optimum. Later, *(2)* extends the analysis to situations where different travelers may have different value of time. *(3)* explores the problem in a network with two parallel routes, combining travelers' departure time choice with route choice. This network configuration is later revisited by *(4)* who find that the route splits of the two parallel routes at user equilibrium and system optimum are identical. *(5)* points out the difficulties in obtaining a realistic measure of welfare across users, and proposes a Pareto-improving tolling strategy in which a time-dependent toll is applied during a time window and some users are exempted from paying it. The classification of road users is such that the fraction of days that a user is free to use the road is the same for all the users in the long run.

Although the idea of using dynamic road tolls to internalize congestion externality is theoretically sound, it is practically difficult to implement. A first-best time-dependent toll has little chance to be accepted by the public because all the links are required to be tolled. Even it is accepted, the sophisticated change in toll charges will inevitably cause confusion among users. Besides, toll booths not only incur construction costs, but also often form new bottlenecks that cause additional delays in an already congested route if not all tolls are paid electronically.

Internalizing externalities by congestion pricing is the direct but NOT the only recipe for improving efficiency. If different groups of travelers' accesses to the bottleneck can be controlled separately, the congestion externality may be limited within the same group of people and thus its impact to the whole system can be alleviated. *(6)*, when analyzing the dynamic user equilibrium traffic pattern in a merge network with tandem bottlenecks, mention the possibility of reducing the total system cost by metering the upstream bottleneck. However, the potential of ramp metering is not fully exploited in that paper because metering rates are assumed to be constant throughout the morning peak. In this paper, we thoroughly explore the idea of *externality redistribution* in a linear freeway with multiple on-ramps and a downstream bottleneck. We propose *Pareto-improving* metering strategies under which no users are worse off. In these metering strategies, each on-ramp has two values of metering rates (low and high). A time window is assigned to each on-ramp within which it operates at its low metering rate and outside which it operates at its high metering rate. By choosing the duration of the time window carefully, these ramps can serve as reservoirs to hold congestion within the same group of people and thus alleviate the interaction of traffic among different groups. The proposed Pareto-improving ramp metering strategies are one type of *imperfect rationing*, but compared to congestion pricing and other rationing strategies, ramp metering makes use of the existing infrastructure of on-ramps, requires relatively low cost to implement and hence is more likely to be accepted by the public.

The rest of the paper is structured as follows. Section 2 describes the problem and Section 3 presents some important properties of the time-dependent user equilibrium traffic pattern in the context of morning commute. These properties are the basis for deriving our Pareto-improving metering plans. Section 4 focuses on a two-ramp freeway network and discusses the idea of using ramp metering to redistribute externalities. Three types of ramp metering plans in descending order of total system cost are presented. Section 5 then extends the results in Section 4 to a more general linear freeway with more than two on-ramps. Finally, Section 6 concludes the paper by summarizing the findings and their practical implications.

2. THE PROBLEM

We consider a linear freeway with m uncapacitated on-ramps and one bottleneck with capacity c at the end of the freeway (Figure 1(a)). The ramps are labeled in ascending order from the one closest to the freeway bottleneck. Meters are installed at all the ramps to regulate freeway entry flow rates. For the ease of presentation, the freeway traversal time under free-flow conditions is assumed to be zero throughout our analysis. It can be shown that by simple modifications, the results derived under the zero free-flow travel time assumption can be easily applied to networks with positive free-flow travel time. With this assumption, the study network can be further simplified as a many-in-one-out merge (Figure 1(b)).

During the morning peak, each ramp i corresponds to an origin O_i, which generates travel demand heading toward the Central Business District(CBD) located at the end of the freeway. The total number of travelers originated from O_i is N_i and the sum of all the travelers in the network is denoted as N, i.e., $N = \sum_{i=1}^{m} N_i$. Travelers using the linear freeway share the same desired arrival time $t^* = \sum_{i=1}^{n} N_i / c$ at the destination. (Note that although the absolute value of t^* does not affect the results in terms of flow distribution, we choose $t^* = \sum_{i=1}^{n} N_i / c$ to make the results easier to present.) For simplicity, the same assumption used in *(6)* - that late arrival is not permitted - is adopted in this study. Travel costs consist of two parts: 1) travel time cost - queueing delays at ramps due to metering and at the freeway bottleneck due to capacity restriction; 2) schedule delay cost - cost associated with arriving at the destination earlier than desired. For any arrival time $t < t^*$ at the destination, the schedule delay cost (converted into travel time units) is assumed to be linearly decreasing in the rate of $\alpha < 1$ *(7)*. The travel cost structure in this study can thus be represented by the following relationships:

$$C_i(t) = T_i(t) + SD[t + T_i(t)]$$
$$T_i(t) = L_i(t) + L[t + L_i(t)]$$
$$SD(t) = \begin{cases} \alpha(t^* - t) & t \le t^* \\ +\infty & \text{otherwise} \end{cases}$$

where

$C_i(t)$ is the total travel cost for a traveler originated from O_i at time t;
$T_i(t)$ is the travel time cost for a traveler originated from O_i at time t;
$L_i(t)$ is the queuing delay at ramp i for a traveler entering the ramp at time t;
$L(t)$ is the queuing delay at the freeway bottleneck for a traveler approaching the bottleneck at time t;
$SD(t)$ is the schedule delay cost for a traveler arriving at the destination at time t.

To facilitate the analysis, denote the cumulative arrival and departure curves at the freeway bottleneck and at ramp i as $A(t)$, $D(t)$, $A_i(t)$ and $D_i(t)$, respectively. Furthermore, the corresponding lower case symbols are used to denote the instant flow rate. For example, $a(t) = dA(t)/dt$ represents the arrival flow rate at the freeway bottleneck at time t while $d(t) = dD(t)/dt$ is the bottleneck departure flow rate.

If no metering plan is implemented, all the travelers were as if from the same origin. According to *(1)*'s morning commute analysis for a single bottleneck, the equilibrium traffic flow pattern in this case can be represented by the cumulative curves at the freeway bottleneck in Figure 2.

As shown, at equilibrium, all the travelers depart within a time period $[0,(1-\alpha)N/c]$ with the aggregate arrival rate at the freeway bottleneck equal to $c/(1-\alpha)$. The freeway bottleneck always operates at its capacity from $t=0$ to $t=t^*$. It is easy to check that the personal cost for any departure time within time period $[0,(1-\alpha)N/c]$ is equal to the equilibrium cost $\bar{C}^0=\alpha t^*=\alpha N/c$ and lower than that for any departure time outside that time window. The total system cost TC^* at equilibrium is therefore $TC^0=N\cdot\alpha N/c=\alpha N^2/c$.

To eliminate the queue at the freeway bottleneck to improve network efficiency, one typical strategy is to internalize travelers' externalities by applying to the freeway bottleneck a time-dependent toll $\tau(t)=\alpha t, t\in[0,t^*]$. Such a tolling scheme requires to enforce continuously changing tolling rates and to install toll booths on either the freeway mainline or entry ramps, which may cause additional delay to travelers.

As an alternative, this study illustrates how time-dependent ramp metering plans can be designed to improve network efficiency based on the idea of *externality redistribution*. To emphasize the applicability in practice, any ramp metering plans proposed in this study, denoted by $\{r_i(t)\}$ with $r_i(t)$ representing the time-dependent metering rate at ramp i at time t, must satisfy the following criteria:

1) To avoid potential confusion to the public, at most two different metering rates (high/low) can be applied to any ramp meter;

2) The lower metering rate at any ramp meter cannot be smaller than a pre-determined value $r_{\min}$ which falls in $[0,cN_i/N],\forall i=1,\ldots,m$; (Note that we require $r_{\min}$ to be less than $cN_i/N, i=1,\ldots,m$ such that if all the meters are operated at $r_{\min}$, not all the travelers can be discharged within a time period with length N/c.)

3) The metering plan is *Pareto-improving*, meaning that the equilibrium cost for travelers from any origin is no higher than the equilibrium cost $\bar{C}^0$ in the no-metering scenario.

3. SOME BASIC FEATURES OF THE USER EQUILIBRIUM TRAFFIC PATTERN IN A LINEAR FREEWAY

This section presents two propositions which summarize some basic features of the user equilibrium traffic pattern in a linear freeway during the morning peak. These two propositions will serve as building blocks for designing Pareto-improving ramp metering plans in the next two sections.

Proposition 1 *At user equilibrium with departure time choice in a linear freeway, for any traveler originated from O_i at time t, the flow rates when she reaches/leaves the ramp and when she reaches/leaves the freeway bottleneck must satisfy the following relationship:*

$$\frac{a_i(t)}{d_i[t+L_i(t)]}\cdot\frac{a[t+L_i(t)]}{d[t+T_i(t)]}=\frac{1}{1-\alpha},\forall t\in\Gamma_i, i=1,\ldots,m \quad (1)$$

where

t *is the time when the traveler arrives at ramp* i *;*

$t+L_i(t)$ *is the time when the traveler leaves ramp* i *and gets to the freeway bottleneck;*

$t+T_i(t)$ *is the time when the traveler departs from the freeway bottleneck;*

Γ_i *is the set of times during which travelers from* O_i *departs.*

Proof. The First-In-First-Out (FIFO) property at the ramps and the freeway bottleneck (Figure 3) implies that

$$A_i(t) = D_i[t + L_i(t)] \quad (2)$$

$$A[t + L_i(t)] = D[t + T_i(t)] \quad (3)$$

Differentiating both (2) and (3) by t, we get

$$\frac{dA_i(t)}{dt} = \frac{dD_i[t + L_i(t)]}{dt}$$

$$\frac{dA[t + L_i(t)]}{dt} = \frac{dD[t + T_i(t)]}{dt}$$

From the Chain rule,

$$\frac{dA_i(t)}{dt} = \frac{dD_i[t + L_i(t)]}{d[t + L_i(t)]}\left(1 + \frac{dL_i(t)}{dt}\right) \quad (4)$$

$$\frac{dA[t + L_i(t)]}{d[t + L_i(t)]}\left(1 + \frac{dL_i(t)}{dt}\right) = \frac{dD[t + T_i(t)]}{d[t + T_i(t)]}\left(1 + \frac{dT_i(t)}{dt}\right) \quad (5)$$

Combining (4) and (5) to eliminate $dL_i(t)/dt$, we get

$$\frac{dT_i(t)}{dt} = \frac{dA_i(t)/dt}{dD_i[t + L_i(t)]/d[t + L_i(t)]} \cdot \frac{dA[t + L_i(t)]/d[t + L_i(t)]}{dD[t + T_i(t)]/d[t + T_i(t)]} - 1$$

Namely,

$$\frac{dT_i(t)}{dt} = \frac{a_i(t)}{d_i[t + L_i(t)]} \cdot \frac{a[t + L_i(t)]}{d[t + T_i(t)]} - 1 \quad (6)$$

Note that relationship (6) does not rely on the equilibrium condition.

On the other hand, we know that the personal travel cost for any traveler departing from O_i at time t is made up of the travel time cost and the schedule delay cost. Namely,

$$C_i(t) = T_i(t) + SD[t + T_i(t)] = T_i(t) + \alpha[t^* - t - T_i(t)] \quad (7)$$

By definition, at equilibrium, $C_i(t)$ must be constant for all $t \in \Gamma_i$ and we denote this equilibrium cost as $\overline{C}_i = C_i(t), \forall t \in \Gamma_i, i = 1, \ldots, m$. Differentiating (7) with respect to *t*, we get

$$\frac{dC_i(t)}{dt} = \frac{dT_i(t)}{dt} - \alpha - \alpha\frac{dT_i(t)}{dt} = 0, \forall t \in \Gamma_i, i = 1, \ldots, m$$

Namely,

$$\frac{dT_i(t)}{dt} = \frac{\alpha}{1-\alpha}, , \forall t \in \Gamma_i, i = 1, \ldots, m \tag{8}$$

(6) and (8) then lead to

$$\frac{a_i(t)}{d_i[t + L_i(t)]} \cdot \frac{a[t + L_i(t)]}{d[t + T_i(t)]} = \frac{1}{1-\alpha}, \forall t \in \Gamma_i, i = 1, \ldots, m \qquad \square$$

Proposition 1 presents an interesting property of the equilibrium traffic pattern in a linear freeway. In particular, if we know three of the four cumulative curves ($A_i(t), D_i(t), A(t), D(t)$) on a traveler's way from her origin to the destination, the remaining one can be constructed easily based on relationship (1). Furthermore, if it is certain that there is to be no congestion at the freeway bottleneck at equilibrium, then $a[t + L_i(t)] = d[t + T_i(t)], \forall t$ and hence $a_i(t) / d_i[t + L_i(t)] = 1/(1-\alpha)$. In these special cases, knowing the arrival curve $A_i(t)$ for ramp i (or departure curve $D_i(t)$) is sufficient for deriving the departure curve $D_i(t)$ (or arrival curve $A_i(t)$) at the same location. When we derive the equilibrium traffic pattern for a certain metering plan in the next two sections, we shall first construct the departure curves both at the ramps and at the freeway bottleneck and then obtain the arrival curves at those locations accordingly based on relationship (1).

Proposition 2 *During the morning peak, if*
1) $\bar{C}_i > \bar{C}_j$ *, i.e., the equilibrium cost of travelers from* O_i *is larger than that of travelers from* O_j *, and*
2) the last traveler from O_j *experiences positive queueing delay on ramp* j *, then,*
the last travelers from O_i *and from* O_j *must reach the freeway bottleneck simultaneously.*

Proof. Suppose t_i and t_j are the times when the last travelers from O_i and O_j arrive at the freeway bottleneck, respectively.

If $t_i < t_j$, the last traveler from O_i can delay her departure and arrive at the freeway bottleneck at time t_j. The personal cost of this traveler is $C_i(t_j) = L(t_j) + SD[t_j + L(t_j)] \le \bar{C}_j < \bar{C}_i$, contradicting with the equilibrium condition.

If $t_i > t_j$, the last traveler from O_j reaches the destination earlier than desired because there are still travelers traversing the freeway bottleneck after she arrives at the destination. Since she experiences positive delay on ramp j, she can postpone her departure till the queue on ramp j dissipates and still arrive at the destination at exactly the same time as the previous one. Evidently, this cost is smaller than $\bar{C}_j$, contradicting with the equilibrium condition.

In summary, $t_i = t_j$. Namely, the last traveler from O_i and from O_j must reach the freeway bottleneck simultaneously. □

In particular, in a linear freeway with only two ramps, Proposition 2 indicates that given conditions 1) and 2), the last travelers from both ramps would reach the bottleneck simultaneously and then arrive at the destination right on time.

4. METERING PLANS IN A TWO-RAMP LINEAR FREEWAY

In this section, we focus on a linear freeway with only two on-ramps. We discuss three metering plans, each of which corresponds to a different externality redistribution strategy, in the order of decreasing total system cost.

4.1 Metering plan I

As stated previously, the basic idea of externality redistribution is to reduce the network-wide congestion impact by localizing traffic interactions within small groups. Obviously, the most straightforward strategy is to restrict the entry flow rates at the ramps so that travelers queue at the ramps instead of at the freeway bottleneck.

According to this idea, the first metering plan we explore maintains constant metering rates proportional to the percentage of travel demand over the total demand for both ramps, i.e.,

Metering plan I: $r_1(t) = c\dfrac{N_1}{N}, \forall t$ and $r_2(t) = c\dfrac{N_2}{N}, \forall t$

Note that both ramps can still discharge all the traffic within time period $[0, N/c]$.

With this metering plan launched, the aggregate arrival flow at the freeway bottleneck never exceeds the bottleneck capacity c, and thus no congestion would occur at the freeway bottleneck. Therefore, the equilibrium traffic patterns for travelers from O_1 and O_2 can each be derived independently in a similar way.

For example, to derive the equilibrium traffic pattern for travelers from O_1 , first notice that the departure rate $d_1(t)$ from ramp 1 should be equal to the metering rate cN_1/N during $t \in [0, t^*]$ where $t^* = N/c$. Otherwise, there must exist a time tick t' such that $d_1(t') < cN_1/N$ and another time tick $t'' < 0$ such that $d_1(t'') > 0$ in order to discharge all the traffic to the freeway. $C_1(t'') > C_1(t')$ because $C_1(t')$ contains only the schedule delay cost, which is obviously lower than that for departure time t''. This conflicts with the equilibrium condition because travelers leaving the ramp at time t'' can delay their departures till time t' to further reduce their costs. Finally, according to relationship (1), $a_1(t) = d_1[t + L_1(t)]/(1-\alpha) = cN_1/[(1-\alpha)N], \forall t \in [0, (1-\alpha)N/c]$. The equilibrium traffic pattern for travelers from O_2 can be attained in the same way. Namely, $d_2(t) = cN_2/N, \forall t \in [0, N/c]$ and $a_2(t) = cN_2/[(1-\alpha)N], \forall t \in [0, (1-\alpha)N/c]$. The equilibrium cumulative curves for the freeway bottleneck and the two ramps are shown in Figure 4.

Evidently, the equilibrium cost for travelers from either origin is equal to the schedule delay cost of the first departing traveler. Namely, $\bar{C}_1^I = \bar{C}_2^I = \alpha N/c$, which is unfortunately still identical to the equilibrium cost $\bar{C}^0$ in the basic scenario. In other words, such a metering plan simply separating the externalties of the two ramps cannot bring benefits to any of the traveler groups.

4.2 Metering plan II

Although metering plan I cannot reduce the total system cost, it does provide some hints for redistributing externalities in a more efficient way. As shown, when no congestion happens at the freeway bottleneck, the equilibrium cost for each traveler group is determined only by the schedule delay of the first departing traveler. This observation indicates that if we give priority to one ramp(say, ramp 2) by increasing its discharging rate r_2 above cN_2/N and reducing the metering rate of the other (ramp 1) at the same time correspondingly, the earliest departure time at the former ramp (ramp 2) can be postponed to a time tick $t_1 = N/c - N_2/r_2$ later than what is in the basic scenario ($t = 0$). By this means, the equilibrium cost for travelers from O_2 can be reduced. In order to guarantee that the equilibrium cost for traveler from O_1 does not increase, the metering rate of ramp 1 before t_1, the earliest departing time on ramp 2, must be increased from cN_1/N to c such that the first traveler departs no earlier than time $t = 0$.

In accord with this externality redistribution idea, the second metering plan we explore involves a constant metering rate $r_2 > cN_2/N$ for ramp 2, and a high metering rate c before time t_1 and a low metering rate $c-r_2$ after time t_1 for ramp 1. Namely,

$$\textbf{Metering plan II:} \quad r_1(t)=\begin{cases} c & t \le \frac{N}{c}-\frac{N_2}{r_2} \\ c-r_2 & t > \frac{N}{c}-\frac{N_2}{r_2} \end{cases} \text{ and } r_2(t)=r_2 > c\frac{N_2}{N}, \forall t$$

Similar to the analysis for metering plan I, the departure rates on both ramps are equal to their metering rates from the time when the first traveler departs to time t^*. Namely,

$$d_1(t)=r_1(t), \forall t \in [0,t^*]$$
$$d_2(t)=r_2(t), \forall t \in [t^*-\frac{N_2}{r_2}, t^*]$$

Since there is still no congestion at the freeway bottleneck in this metering plan, $a[t]=d[t+L(t)], \forall t$. Relationship (1) then leads to $a_i(t)=d_i[t+L_i(t)]/(1-\alpha), i=1,2$. By simple calculations, we can get

$$a_1(t)=\begin{cases} \frac{c}{1-\alpha} & t \in [0,(1-\alpha)(t^*-\frac{N_2}{r_2})] \\ \frac{c-r_2}{1-\alpha} & t \in (t^*-\frac{N_2}{r_2},(1-\alpha)\frac{N}{c}] \\ 0 & \text{otherwise} \end{cases}$$
$$a_2(t)=\begin{cases} \frac{r_2}{1-\alpha} & t \in [t^*-\frac{N_2}{r_2}, t^*-\alpha\frac{N_2}{r_2}] \\ 0 & \text{otherwise} \end{cases}$$

The equilibrium cumulative curves at the freeway bottleneck and at the two ramps for this metering plan are shown in Figure 5. The equilibrium pattern for metering plan I is also depicted as the light dash lines for comparison.

In this new equilibrium traffic pattern, the equilibrium cost for travelers from O_1 is $\bar{C}_1^{II}=\alpha N/c$ which is the same as $\bar{C}^0$ in the basic scenario while that for travelers from O_2 is reduced from $\bar{C}^0$ to $\bar{C}_2^{II}=\alpha N_2/r_2$ since $r_2 > cN_2/N$. Therefore, metering plan II is indeed a Pareto-improving plan.

Under metering plan II, $\bar{C}_2^{II}$ is decreasing as r_2 gets larger. The maximal value of r_2 is $c-r_{\min}$ since the metering rate of ramp 1, which is $c-r_2$ during time $[t_1,0]$, cannot be below $r_{\min}$. Therefore, the minimal $\bar{C}_2^{II}$ for this type of metering plan is $\alpha N_2/(c-r_{\min})$ and we call the corresponding plan *Metering plan II**. Namely,

$$\textbf{Metering plan II}^{*}\textbf{:} \quad r_1(t)=\begin{cases} c & t \le \frac{N}{c}-\frac{N_2}{c-r_{\min}} \\ r_{\min} & t > \frac{N}{c}-\frac{N_2}{c-r_{\min}} \end{cases} \text{ and } r_2(t)=c-r_{\min}, \forall t$$

Figure 6 shows the equilibrium traffic pattern for metering plan II* with $r_2=c-r_{\min}$. In the very extreme case where $r_{\min}$ can be as small as zero (Figure 6(a.2) - (c.2)), the duration time when travelers from O_1 and O_2 traverse the freeway can be completely separated and $\bar{C}_2^{II}$ can be as small as

$\alpha N_2 / c$.

4.3 Metering plan III

In both metering plans I and II, no congestion is present at the freeway bottleneck throughout the morning peak. We now continue to explore if network performance can be further improved by allowing queues at the freeway bottleneck within a certain time window.

In metering plan II, at the second stage of the morning peak when travelers from both origins exist in the network, if we further give more priority to ramp 2 by raise its metering rate r_2 above $c - r_{\min}$ while keeping the metering rate of ramp 1 unchanged (i.e., at $r_{\min}$), queues would develop not only at the ramps but also at the freeway bottleneck. Such a strategy enables the first traveler from O_2 to further delay her departure to a later time $t_1 > N / c - N_2 /(c - r_{\min})$. At the same time, because of the presence of queues at the freeway bottleneck, the last travelers from both ramp 1 and ramp 2 must arrive at the freeway bottleneck earlier than time t^* in order to arrive at the destination on time. This idea leads to a new metering plan which maintains constant metering rate $r_2 > c - r_{\min}$ for ramp 2, and a high metering rate c before t_1 and a low metering rate $r_{\min}$ after t_1 for ramp 1.

To complete this new metering plan, we still need to determine t_1 which is the time tick separating the high and low metering rate for ramp 1 and also the departure time of the first traveler from O_2. Note that according to Proposition 2, the last travelers from ramp 1 and ramp 2 arrive at the freeway bottleneck at the same time. Let this time be t'. Since ramp 2 will discharge flows equal to r_2 during $[t_1, t']$, we have

$$t_1 = t' - N_2 / r_2 \tag{9}$$

On the other hand, $t^* - t'$ is the freeway queuing delay that the last travelers from both ramps experience and thus can be determined by the queuing relationship at the freeway bottleneck. Namely,

$$t^* - t' = (r_2 + r_{\min} - c) \cdot \frac{N_2}{r_2} / c \tag{10}$$

Combining (9) and (10) to solve for t' and t_1, we get

$$t' = \frac{N_1}{c} + \frac{N_2}{r_2} \frac{c - r_{\min}}{r_2} \tag{11}$$

$$t_1 = \frac{N_1}{c} - \frac{N_2}{r_2} \frac{r_{\min}}{c} \tag{12}$$

Therefore, the third metering plan we explore is as follows:

Metering plan III: $r_1(t) = \begin{cases} c & t \le \frac{N_1}{c} - \frac{N_2}{r_2} \frac{r_{\min}}{c} \\ r_{\min} & t > \frac{N_1}{c} - \frac{N_2}{r_2} \frac{r_{\min}}{c} \end{cases}$ and $r_2(t) = r_2 > c - r_{\min} \forall t$

Before we derive the equilibrium traffic pattern for metering plan III, we need to first check whether this metering plan would increase the equilibrium cost for travelers from O_1. Let the time when the first traveler from O_1 departs be t''. Since ramp 1 always discharges flows equal to its metering rate from t'' to t', we have

$$(t_1 - t'')c + (t' - t_1)r_{\min} = N_1 \tag{13}$$

Substituting the values of t_1 and t' from (11) and (12) into (13), we get $t''=0$, meaning that the equilibrium cost for travelers from O_1, which is equal to the schedule delay of the first departing traveler, is exactly the same as what is in the basic scenario. Hence, metering plan III is also a Pareto-improving plan.

We have now derived the departure curves for both ramps. The arrival curve and the departure curve at the freeway bottleneck are easy to get since the former one is just the sum of the departure curves of the two ramps and the latter one is a straight line with slope c during $[0,t^*]$. Now we can apply (1) to derive the arrival curves at both ramps.

For ramp 1, we have

$$a_1(t) = \begin{cases} \frac{d_1[t+L_1(t)]}{1-\alpha} = \frac{c}{1-\alpha} & t \in [0, t_1^1] \\ \frac{d_1[t+L_1(t)]}{1-\alpha} \frac{d[t+T(t)]}{a[t+L_1(t)]} = \frac{r_{\min}}{1-\alpha} \frac{c}{r_2 + r_{\min}} & t \in (t_1^1, t_1^2] \\ 0 & \text{otherwise} \end{cases}$$

where t_1^1 and t_1^2 are the time ticks when the arrival rate on ramp 1 changes. According to the flow conservation relationship,

$$t_1^1 = \frac{(t_1 - 0)c}{a_1(t)} = (1-\alpha)\left[\frac{N_1}{c} - \frac{r_{\min} N_2}{c r_2}\right], \text{ and } t_1^2 = \frac{(t' - t_1) r_{\min}}{a_1(t)} = (1-\alpha)\frac{N}{c}$$

For ramp 2, we have

$$a_2(t) = \begin{cases} \frac{d_2[t+L_2(t)]}{1-\alpha} \frac{d[t+T(t)]}{a[t+L_2(t)]} = \frac{r_2}{1-\alpha} \frac{c}{r_2 + r_{\min}} & t \in [\frac{N_1}{c} - \frac{N_2}{c}\frac{r_{\min}}{r_2}, t_2^1] \\ 0 \text{ otherwise} \end{cases}$$

where t_2^1 is the time tick when the arrival rate on ramp 2 becomes zero. According to the flow conservation relationship,

$$t_2^1 = \frac{(t' - t_1) r_2}{a_2(t)} = \frac{N}{c} - \frac{\alpha N_2}{c} \frac{r_{\min} + r_2}{r_2}$$

The equilibrium traffic pattern for metering plan III is shown in Figure 7. The light dash line denotes the equilibrium pattern for metering plan II with $r_2 = r_{\min}$ for comparison.

As shown, the equilibrium cost $\bar{C}_2^{III}$ for travelers from O_2 is

$$\bar{C}_2^{III} = \alpha(t^* - t_1) = \frac{\alpha N_2}{c}\left(1 + \frac{r_{\min}}{r_2}\right)$$

which is decreasing as r_2 gets larger. However, r_2 cannot goes to infinite because at a certain value of r_2, there would be no queue on ramp 2 and the equilibrium cost C_2^{III} reaches its minimum. After that, even if we continue to increase r_2, the equilibrium traffic pattern would not change anymore. In this case, $a_2(t) = d_2(t)$, i.e., $\frac{c}{1-\alpha}\frac{r_2}{r_{\min}+r_2} = r_2 \Rightarrow r_2 = \frac{c}{1-\alpha} - r_{\min}$ and $C_2^{II} = \frac{\alpha N_2}{c-(1-\alpha)r_{\min}}$. We call the corresponding plan *Metering plan III** which can be stated as follows:

Metering plan III*: $r_1(t) = \begin{cases} c & t \le \frac{N_1}{c} - \frac{N_2}{r_2}\frac{r_{\min}(1-\alpha)}{c-(1-\alpha)r_{\min}} \\ r_{\min} & t > \frac{N_1}{c} - \frac{N_2}{r_2}\frac{r_{\min}(1-\alpha)}{c-(1-\alpha)r_{\min}} \end{cases}$ and $r_2(t) = r_2 \ge \dfrac{c}{1-\alpha} - r_{\min} \forall t$ or not metered

The equilibrium traffic pattern for this metering plan III* is shown in Figure 8.

4.4 Summary

By far, we have designed three types of metering plans for a two-ramp freeway. In all these metering plans, one of the ramp (always ramp 2 in our analysis) is metered with constant metering rate (or not metered) while the other one (ramp 1) has two metering rates(high/low). The equilibrium cost of the travelers from O_1 is constant and the same as what is in the basic scenario, while the equilibrium cost of the travelers from O_2 is decreasing as the constant metering rate r_2 gets larger. When r_2 exceeds $c/(1-\alpha) - r_{\min}$, $\bar{C}_2$ reaches its minimum value $\alpha N_2/[c-(1-\alpha)r_{\min}]$ and remains constant. The relationship between r_2 and the equilibrium cost $\bar{C}_2$ can be represented by the solid line in Figure 9.

The difference between the metering plans indicates that simply eliminating the interactions of traffic at the freeway bottleneck by setting each ramp a metering rate proportional to their demand percentage (Metering plan I) is not beneficial for the network. To reduce the system cost, it is necessary to *prioritize* the traveler groups and *partially separate their temporal usage* of the freeway bottleneck. More specifically, the *de-prioritized* ramp (i.e., ramp 1) need to maintain a high metering rate during the first part of the morning peak and a low metering rate during the second part. By this means, travelers from the *prioritized* ramp only need to traverse the freeway bottleneck during the second part of the morning peak and this leads to a lower equilibrium cost for them than what is in the basic scenario. Travelers from the de-prioritized ramp use the freeway bottleneck during both parts of the morning peak and their equilibrium cost is the same as what is in the basic scenario.

Interestingly, the analysis also shows that it is not always beneficial to completely eliminate congestion at the freeway bottleneck. During the first part of the morning peak when only travelers from the de-proritized ramp use the freeway bottleneck, congestion should be held at the *de-prioritized* ramp instead of at the freeway bottleneck so that the first traveler from the prioritized group would not experience queuing delay when she enters the freeway. However, during the second part of the morning peak when travelers from both ramps use the freeway bottleneck, releasing the queues from the *prioritized* ramp to the freeway bottleneck is actually beneficial for the system because it would make travelers from the de-prioritized ramp favor less the second part of the morning peak period and thus give higher priority to travelers from the prioritized ramp.

To implement the Pareto-improving metering plans in practice, we also need to determine which ramp to prioritize. If the total system cost is the only concern, we may want to select a prioritized ramp which results in a lower system cost. Suppose metering plan III*, which results in the minimum system cost among the three types of metering plans we discussed, is the metering plan we want to implement. The total cost reduced from the basic scenario if ramps 1 and 2 are prioritized are denoted as RC_1 and RC_2, respectively. According to our analysis,

$$RC_1 = N_1(\bar{C}_1^{III} - \bar{C}^0) = N_1(\frac{\alpha N}{c} - \frac{\alpha N_1}{c-(1-\alpha)r_{\min}}) = -\frac{\alpha N_1^2}{c-(1-\alpha)r_{\min}} + \frac{\alpha N}{c} N_1$$

$$RC_2 = -\frac{\alpha N_2^2}{c-(1-\alpha)r_{\min}} + \frac{\alpha N}{c} N_2$$

Therefore,

$$\begin{aligned} RC_2 - RC_1 &= -\frac{\alpha}{c-(1-\alpha)r_{\min}}(N_2 - N_1)N + \frac{\alpha N}{c}(N_2 - N_1) \\ &= \alpha N (N_2 - N_1)\left(\frac{1}{c} - \frac{1}{c-(1-\alpha)r_{\min}}\right) \end{aligned} \quad (14)$$

As shown from (14), if $r_{\min} = 0$, i.e., the lower metering rate can be as small as zero, then $1/c - 1/[c-(1-\alpha)r_{\min}] = 0$ and thus $RC_2 = RC_1$; Similarly, if $\alpha = 1$ meaning that the unit cost of early arrival is the same as the unit travel cost, we also have $RC_2 = RC_1$. In both cases, metering either of the two ramps would generate the same system improvement.

In a more general case where $\alpha < 1$ and $r_{\min} > 0$, then $\frac{1}{c} - \frac{1}{c-(1-\alpha)r_{\min}} < 0$. Therefore, $RC_2 > RC_1$ if $N_2 < N_1$ and $RC_2 < RC_1$ if $N_2 > N_1$. Hence, to achieve a lower system cost, one should always prioritize the ramp which has a lower demand.

Finally, if equity is an important concern, another implementation option may be to switch the ramp to be prioritized periodically so that both traveler groups share the benefits from the Pareto-improving metering plan in the long run.

5. PARETO-IMPROVING METERING PLAN IN A MULTI-RAMP LINEAR FREEWAY

This idea of *externality redistribution by prioritizing the ramps* can also be used to design Pareto-improving metering plans for linear freeways with more than two on-ramps. Although there may exists many different ways to prioritize the ramps, we present one type of Pareto-improving metering plan which makes use of the externality redistribution ideas we covered in the previous section.

To design the Pareto-improving metering plan for the multi-ramp linear freeway, a priority order is first decided for all the ramps. As we stated previously, this priority order can be changed for the ramps periodically so that all the traveler groups in the network can benefit from the metering plan. For illustration purpose, we assume that the ramps are already indexed in ascending order of priority from 1 to m. In this Pareto-improving metering plan, each ramp *i* operates at a high metering rate before time t_i and a low metering rate equal to $r_{\min}$ after t_i. The entire metering plan can be designed in the following two steps:

Step I: Prioritize ramp m over the rest ramps

In this step, we design a metering plan III* for the linear network containing ramp *m* and an 'aggregate' ramp made up of ramp *1* through *m-1*, using ramp *m* as the prioritized ramp and the 'aggregate' ramp as the de-prioritized ramp. Note that the minimal metering rate for the 'aggregate' ramp is $(m-1)r_{\min}$ instead of $r_{\min}$. The metering plan III* thus provides the equilibrium arrival and departure curves at ramp *m* and at the freeway bottleneck, as well as the equilibrium departure curve at the 'aggregate' ramp (Figure 10).

As shown, ramp *m* has no congestion and does not need to be metered. Therefore, we can set

$$t_m = \frac{\sum_{i=1}^{m-1} N_i}{c} + \frac{N_m[c-(m-1)r_{\min}](1-\alpha)}{c[c-(1-\alpha)(m-1)r_{\min}]}$$

which is the time when the last traveler from O_m leaves ramp *m*. The high metering rate of ramp *m* can be set as $+\infty$.

The time when ramp *m* begins to have flows, $\frac{\sum_{i=1}^{m-1} N_i}{c} - \frac{N_m}{c}\frac{(m-1)(1-\alpha)r_{\min}}{c-(m-1)(1-\alpha)r_{\min}}$, is the time when the 'aggregate' ramp starts to operate at the low metering rate. We set

$$t_{m-1} = \frac{\sum_{i=1}^{m-1} N_i}{c} - \frac{N_m}{c}\frac{(m-1)(1-\alpha)r_{\min}}{c-(m-1)(1-\alpha)r_{\min}}$$

Since after time t_{m-1}, each ramp $i = 1,\dots,m-1$ only discharges traffic till time t_m, the total number of travelers released before time t_{m-1}, denoted as N_i^1, can be calculated as $N_i^1 = N_i - r_{\min}(t_m - t_{m-1})$. $\{N_i^1\}_{i=1,\dots,m-1}$ will be used in the next step for prioritizing ramps *1* through *m-1*.

The equilibrium departure curve at the 'aggregate' ramp indicates that the sum of the metering rates at ramp *1* through *m-1* should satisfy the following relationship:

$$\sum_{i=1}^{m-1} r_i(t) = \begin{cases} c & t \le t_{m-1} \\ (m-1)r_{\min} & t > t_{m-1} \end{cases} \tag{15}$$

According to the analysis of metering plan III* , if $\sum_{i=1}^{m-1} r_i(t)$ satisfies the above relationship, the equilibrium cost of travelers from O_m would be $\bar{C}_m = \alpha(t^* - t_m)$ which is smaller than $\bar{C}_0$. The equilibrium cost of the travelers from the rest origins O_1 through O_{m-1} now still remain $\bar{C}_0$ and we shall show how to reduce their costs in the next step.

Step II: Prioritize ramp 1, , m – 1

The time t_i separating the high and low metering rates and the value of the high metering rate for ramp $i = 1,\dots,m-1$ can be determined in an iterative manner.

During each iteration ν from $\nu = 1$ to $\nu = m-2$, we prioritize ramp $m-\nu$ over an 'aggregate' ramp made up of ramp 1 through $m-\nu-1$. Each iteration ν determines

1) the high metering rate for ramp $m-\nu$;

2) the time $t_{m-\nu-1}$ when ramp $m-\nu-1$ starts to operate at the low metering rate, which is also the first departure time of travelers from $O_{m-\nu}$;

3) the total number of travelers released before time $t_{m-\nu+1}$ for each ramp $i = 1,\dots,m-\nu-1$. (denoted by $N_i^{\nu+1}$)

For example, when $\nu = 1$, relationship (15) indicates that the sum of the metering rates for ramps *1* through *m-1* before time t_{m-1} should be equal to c. The idea of metering plan II* can then be utilized to prioritize ramp $m-1$ over an 'aggregate' ramp made up of ramp *1* through *m-2* . More specifically, let ramp *m-1* always discharge at the high metering rate $c-(m-2)r_{\min}$ before time t_{m-1} with the earliest departure time $t_{m-2} = t_{m-1} - \frac{N_{m-1}^1}{c-(m-2)r_{\min}}$ and the 'aggregate' ramp discharge at c before time t_{m-2} and at $(m-2)r_{\min}$ from time t_{m-2} to time t_{m-1}. The departure curves at ramp m-1 and at the 'aggregate' ramp are shown in Figure 11.

By this means, the equilibrium cost of travelers from O_{m-1} is reduced from $\bar{C}^0$ in the basic scenario to $\bar{C}_{m-1} = \alpha(t^* - t_{m-2})$.

The sum of the metering rates of ramps *1* through *m-2* should now satisfy the following relationship:

$$\sum_{i=1}^{m-2} r_i(t) = \begin{cases} c & t \le t_{m-2} \\ (m-2) r_{\min} & t > t_{m-2} \end{cases} \tag{16}$$

and the total number of travelers discharged before time t_{m-2} for each ramp $i = 1, \ldots, m-2$ is $N_i^2 = N_i - r_{\min}(t_m - t_{m-2})$.

The same idea can be applied to further prioritize ramp $i = m-2, m-3, \ldots, 2$ iteratively. By the end of the second step, we obtain $t_{m-2}, t_{m-3}, \ldots, t_1$ and the high metering rate for each ramp $i = 1, \ldots, m-1$ is $c - (i-1) r_{\min}$. The departure curves on each ramp are shown in Figure 12.

Once the metering plan is constructed, Proposition 1 can be utilized to determine the arrival curve on each ramp $i = 1, \ldots, m-1$. Figure 13 illustrates both the arrival curve and departure curve on a typical ramp *i*.

Under this metering plan, the equilibrium cost for travelers from each ramp is

$$\bar{C}_i = \begin{cases} \alpha t^* = \bar{C}^0 & i = 1 \\ \alpha (t^* - t_{i-1}) < \bar{C}^0 & i = 2, \ldots, m \end{cases}$$

Apparently, it is indeed a Pareto-improving metering plan.

6. CONCLUDING REMARKS

The morning commute problem has been studied for decades, which led to a better understanding of the congestion formation-and-dissipation phenomenon in traffic networks and corresponding pricing schemes to alleviate congestion. Though congestion pricing is theoretically sound and generates a revenue stream, it also incurs high construction costs, potentially adds delay at toll booths, and is met sometimes with public resistance. In this paper, we propose an alternative approach, ramp metering, to ration the access to the freeway bottleneck by different traveling groups, so as to restrict the congestion externality to within each traveling group, thus reducing peak period congestion. The central idea of our approach is *externality redistribution* as opposed to *externality internalization* in congestion pricing. Basically, by redistributing the externalities originally concentrated all at the freeway bottleneck in the no-metering scenario to different on-ramps via metering, congestion is localized and the total system cost can be reduced. Although ramp metering cannot eliminate all the congestion in the network, hence is second-best, it is much easier and cheaper to implement and involves lower operating cost compared to the first-best marginal-cost congestion pricing.

Unlike conventional metering schemes, our metering strategies are designed to ration the access to the freeway bottleneck by first setting a priority order among the metered ramps then temporally separating their usage of the bottleneck according to their priority. Each ramp, except the ramp with the highest priority, has two metering periods: in the first period a ramp is metered at a high rate and in the second period the low rate. Ramps with lower priority are metered more heavily in the second period (i.e., closer to work start time), so that travelers from high priority ramps can depart at a later time and still experience less travel delay. Interestingly, it is not always harmful to have some congestion at the freeway bottleneck. Instead, releasing the queues from the ramp with the highest priority to the freeway bottleneck can benefit the system, because the presence of congestion at the freeway bottleneck at times close to the work start time would discourage travelers from other ramps to travel during those times.

As compared to most ramp metering strategies implemented in practice, the metering plans proposed in this paper have several attractive features: 1) they take into account travelers' departure time choice, hence capture the effects of metering on peak-spreading; 2) they are all Pareto-improving in the sense that no travelers in the network are worse off compared to no metering; 3) the metering rates for each ramp change at most once during the morning peak so that they can be advertised to travelers ahead of time. Such consistency in the metering plan can reduce day-to-day fluctuations in traffic demand at the metered ramps; and finally 4) they require relatively few data - only the total number of travelers taking each on-ramp.

It is interesting to note that a metering of this kind is in operation in Shanghai, China, where certain entry ramps on the inner ring road are closed during certain hours to reduce the demand pressure on downstream bottlenecks. It is not known to us, however, if the closure times of these ramps are optimized as was done in this paper.

Finally, we want to note that the network we studied here, a single freeway with only entry ramps, is somewhat idealized. Nevertheless, the insights derived in this paper may be applicable to design Pareto-improving metering plans for general networks. It would be quite interesting and challenging to study this latter problem and we leave it to our future work.

ACKNOWLEDGEMENTS

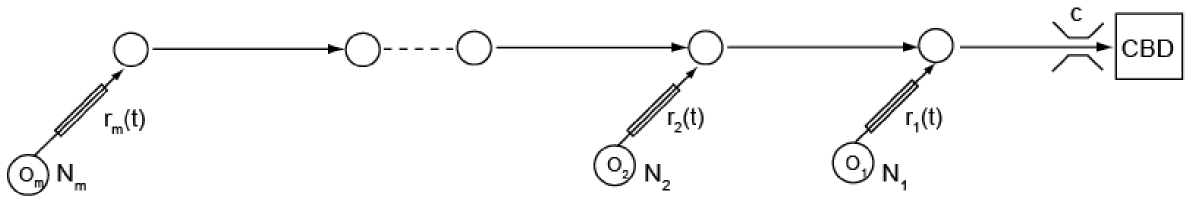
This research is supported by a faculty grant R01-4 from the Sustainable Transportation Center at UCDavis. The views are those of the authors alone.

REFERENCES

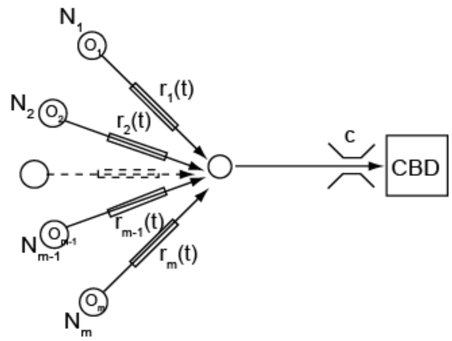
1. Vickrey, W. S.. Congestion Theory and Transport Investment. *American Economic Association* Vol. 59, 1969, pp. 251-261
2. Newell, G. F. The Morning Commute for Nonidentical Travelers. *Transportation Science* Vol. 21(2), 1987, pp. 74-88
3. Mahmassani, H. and R. Herman. Dynamic User Equilibrium Departure Time and Route Choice on Idealized Traffic Arterials. *Transportation Science* Vol. 18(4), 1984, pp. 362-384
4. Arnott, R., A. De Palma, and R. Lindsey. Departure Time and Route Choice for the Morning Commute. *Transportation Research, Part B* Vol. 23(3), 1990, pp. 209-228
5. Daganzo, C. F., and R. C. Garcia. A Pareto Improving Strategy for the Time-Dependent Morning Commute Problem. *Transportation Science* Vol. 34(3), 2000, pp. 303-311
6. Arnott, R., A. De Palma, and R. Lindsey. Properties of Dynamic Traffic Equilibrium Involving Bottlenecks, Including a Paradox and Metering. *Transportation Science* Vol. 27(2), 1993, pp. 148-160
7. Small, K. The Scheduling of Consumer Activities: Work Trips. *American Economic Review* Vol. 72, 1982, pp. 467-479

LIST OF TABLES AND FIGURES

FIGURE 1 The study network
FIGURE 2 The equilibrium traffic flow pattern at the freeway bottleneck with no metering plan
FIGURE 3 An illustration of the FIFO property at ramp i and at the freeway bottleneck
FIGURE 4 The cumulative curves for metering plan I
FIGURE 5 The cumulative curves for metering plan II
FIGURE 6 The cumulative curves for metering plan II* with the minimal $\overline{C}_2^{II}$
FIGURE 7 The cumulative curves for metering plan III
FIGURE 8 The cumulative curves for metering plan III with the minimal $\overline{C}_2^{III}$
FIGURE 9 The relationship between r_2 and $\overline{C}_2$
FIGURE 10 The cumulative curves: step I
FIGURE 11 The cumulative curves: step II
FIGURE 12 The departure curves on each ramp
FIGURE 13 The departure and arrival curves on ramp i



(a) The original network



(b) The abstract study network

FIGURE 1 The study network

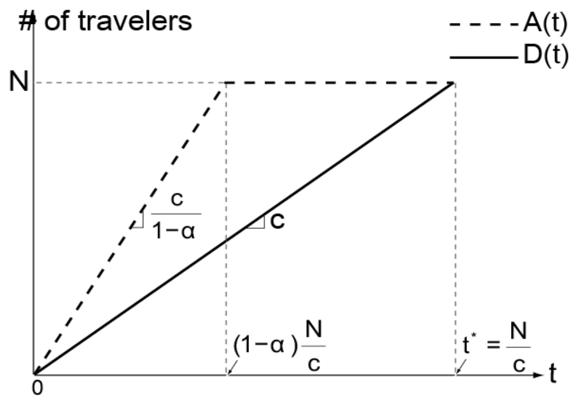
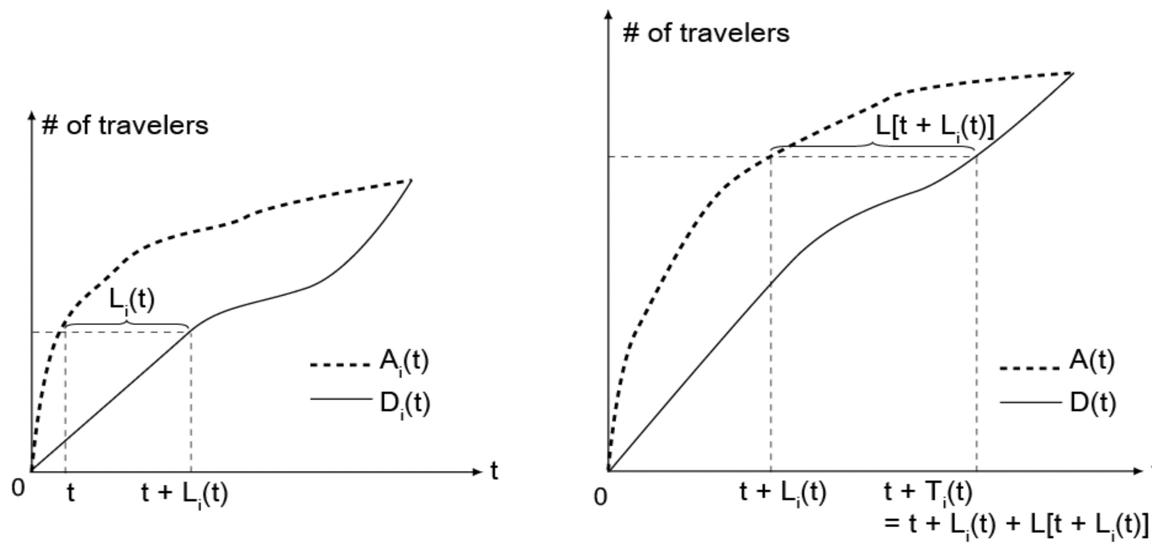


FIGURE 2 The equilibrium traffic flow pattern at the freeway bottleneck with no metering plan



(a) Ramp i (b) Freeway bottleneck

FIGURE 3 An illustration of the FIFO property at ramp i and at the freeway bottleneck

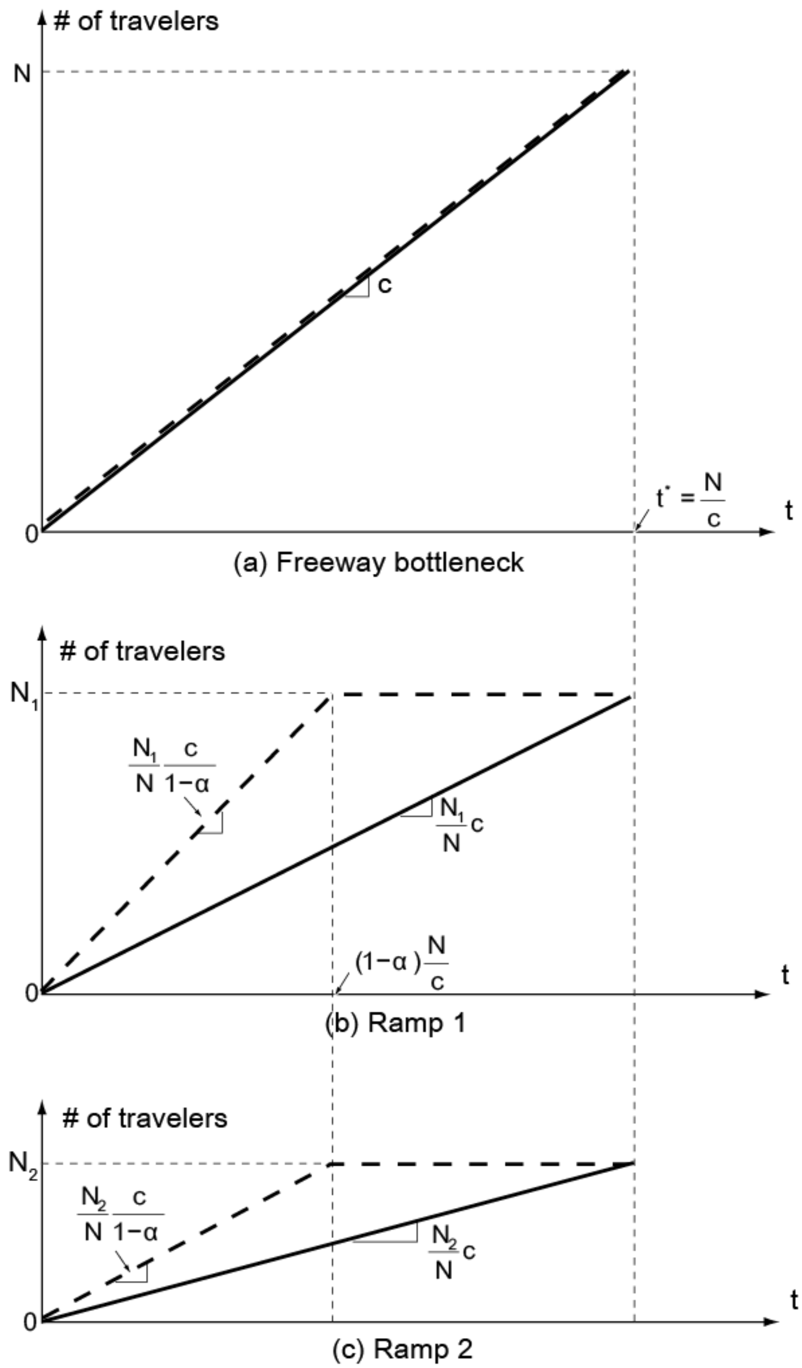


FIGURE 4 The cumulative curves for metering plan I

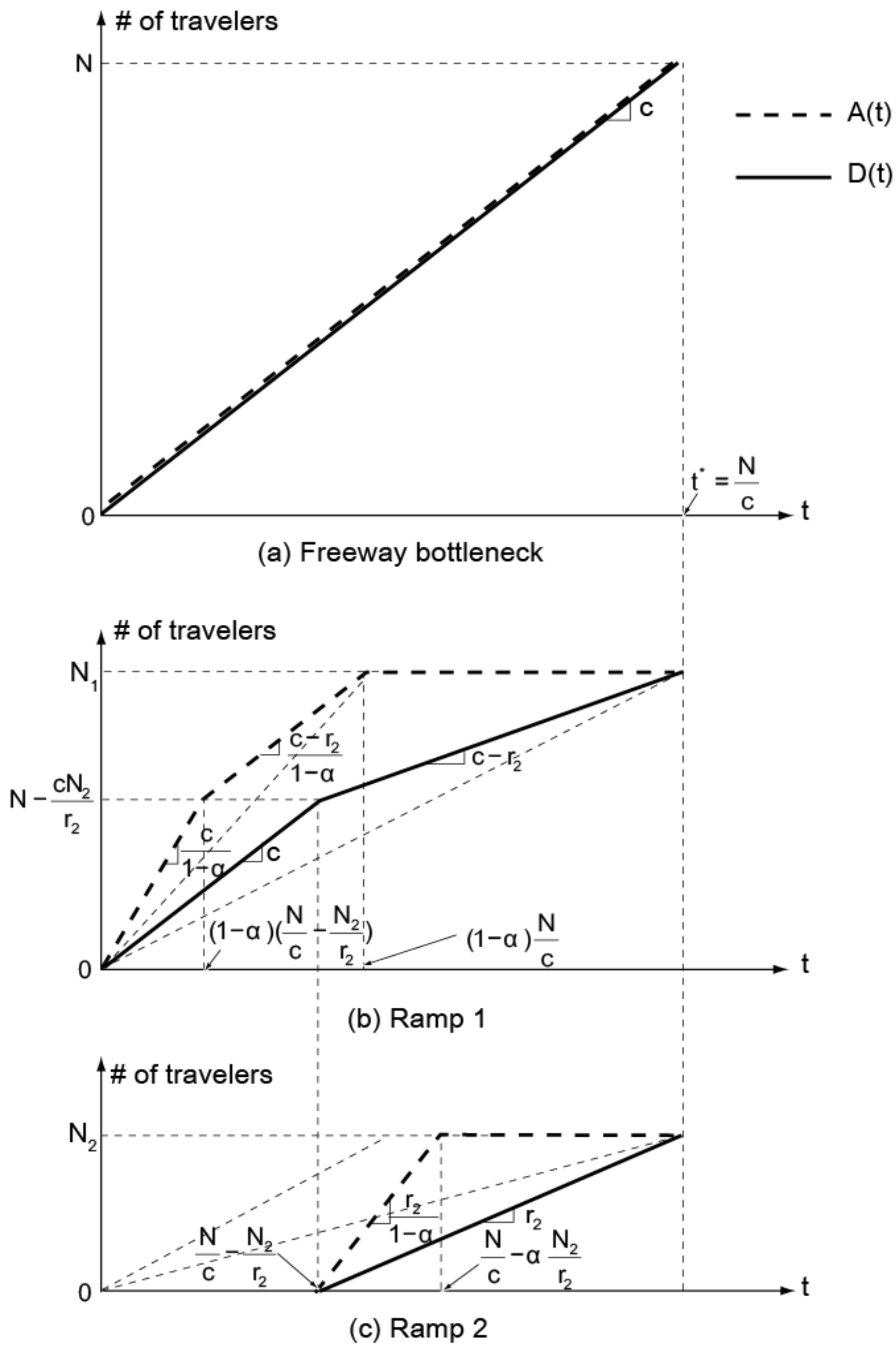


FIGURE 5 The cumulative curves for metering plan II

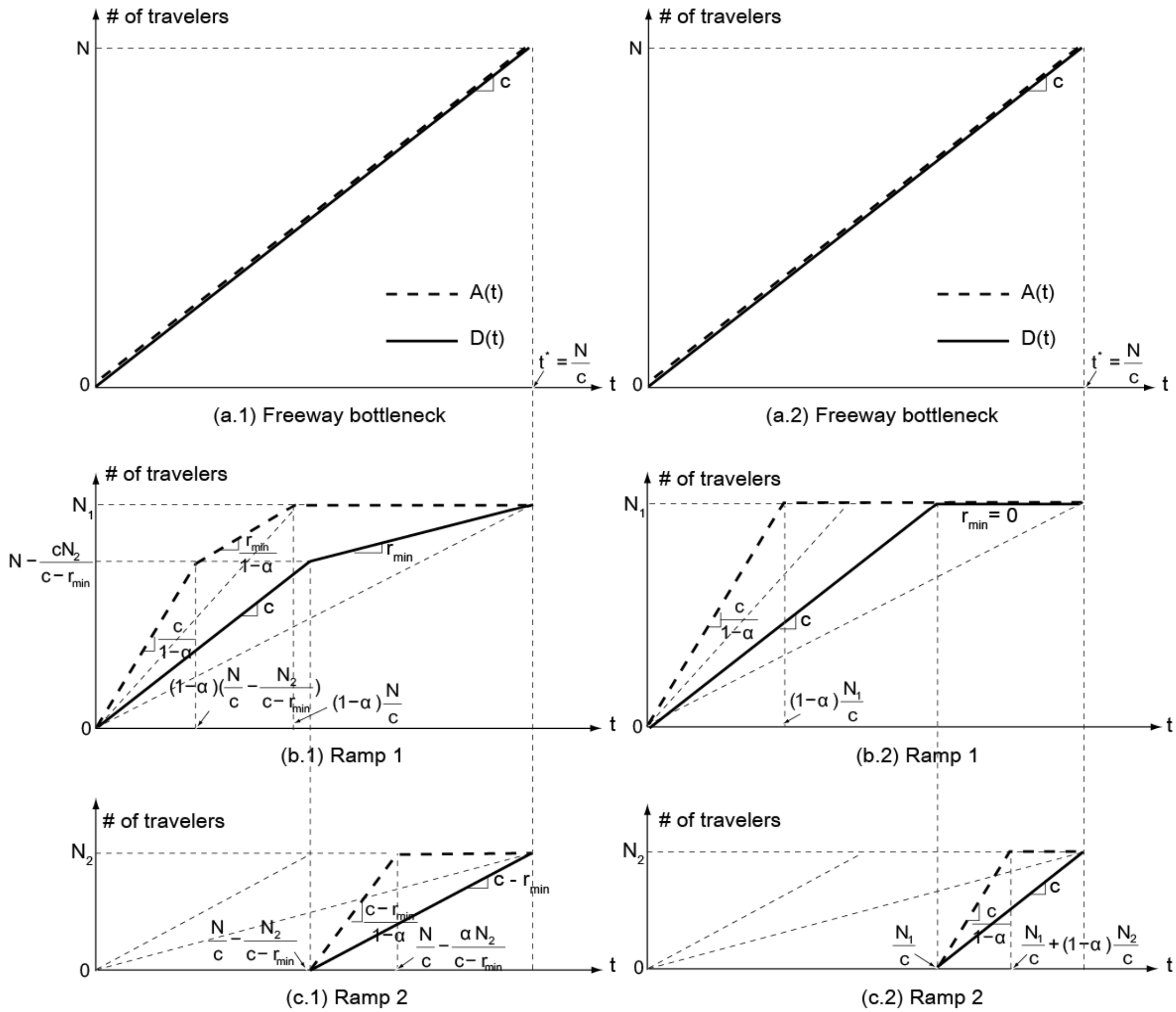


FIGURE 6 The cumulative curves for metering plan II* with the minimal $\overline{C}_2^{II}$

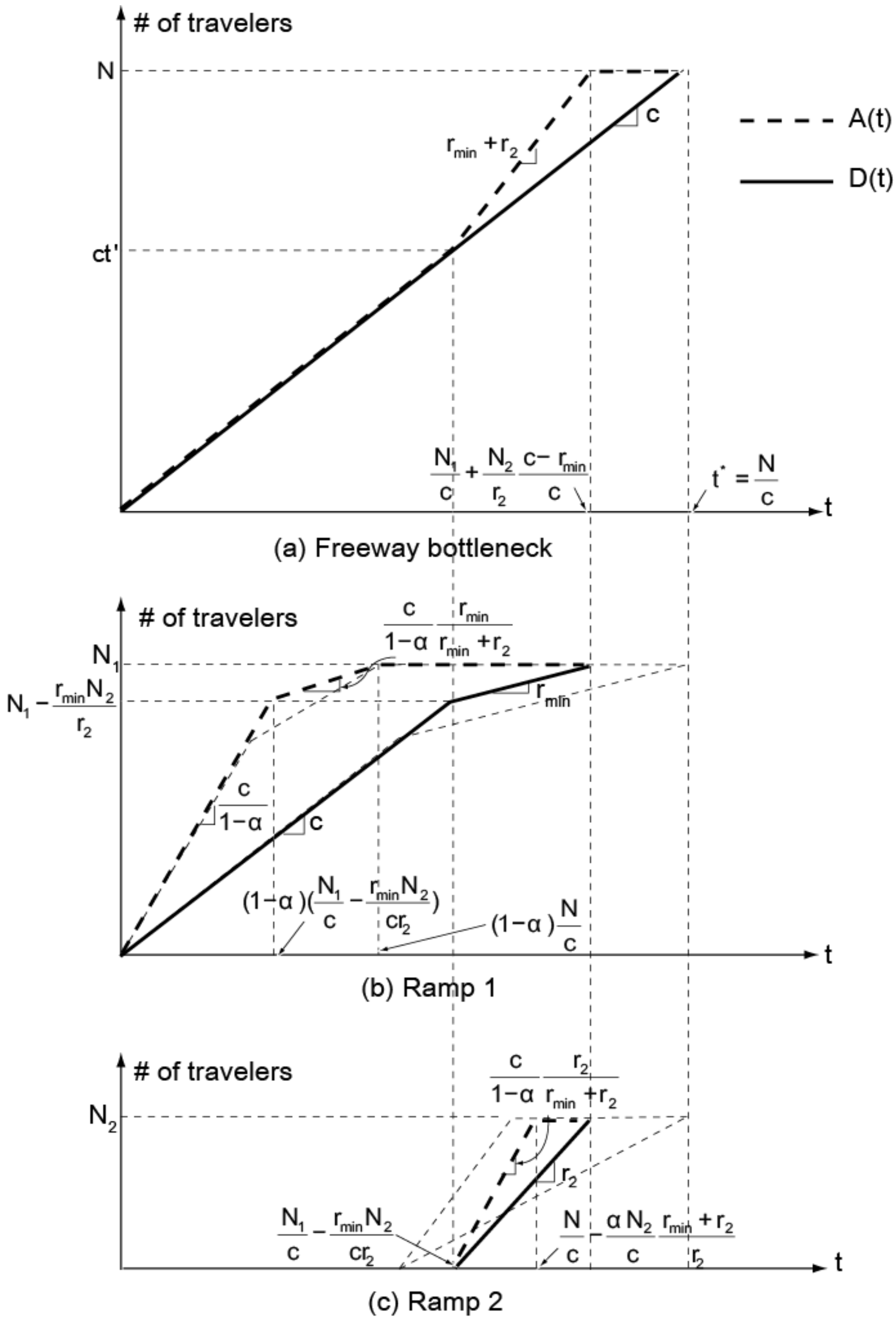


FIGURE 7 The cumulative curves for metering plan III

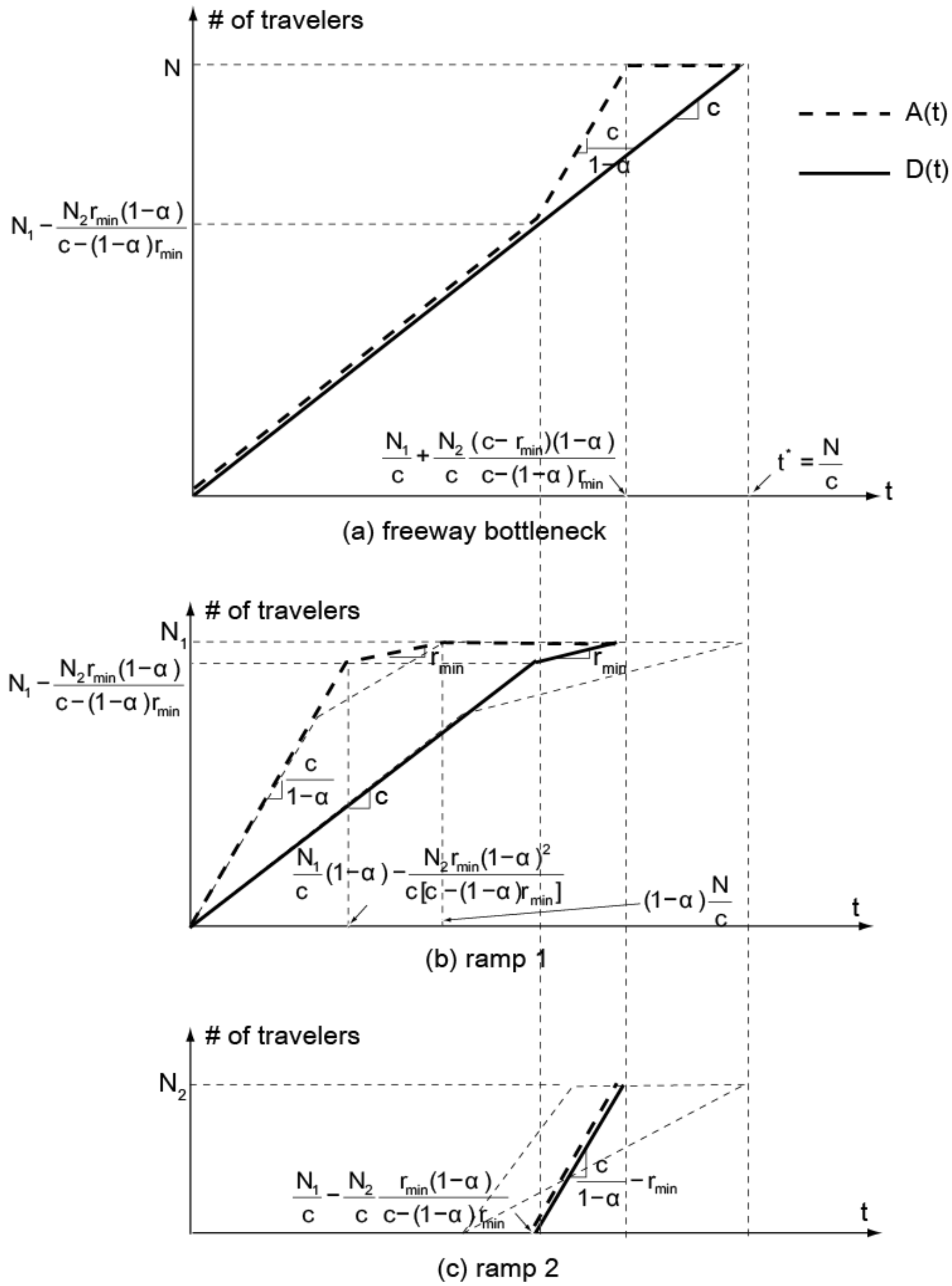


FIGURE 8 The cumulative curves for metering plan III with the minimal $\overline{C}_2^{III}$

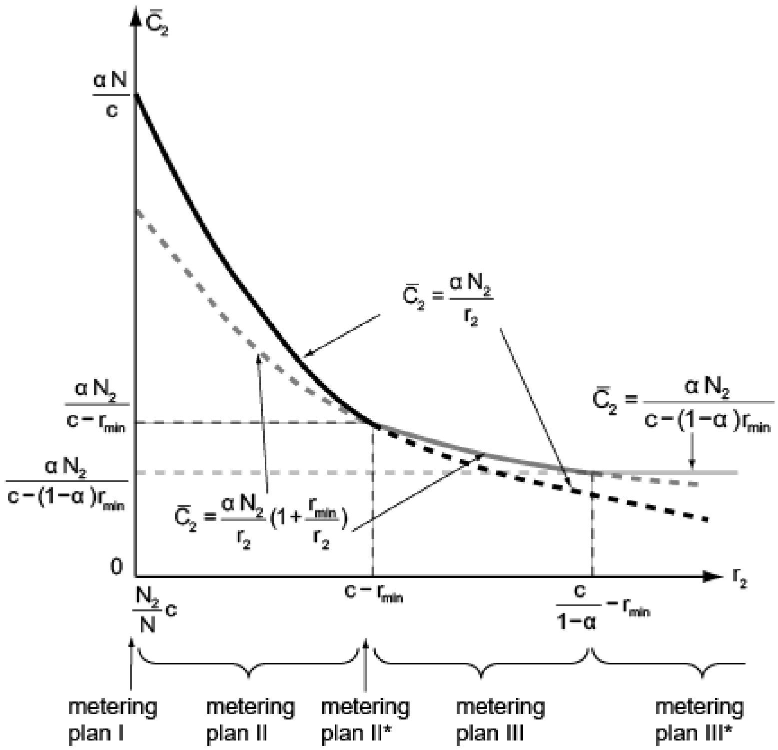


FIGURE 9 The relationship between r_2 and $\overline{C}_2$

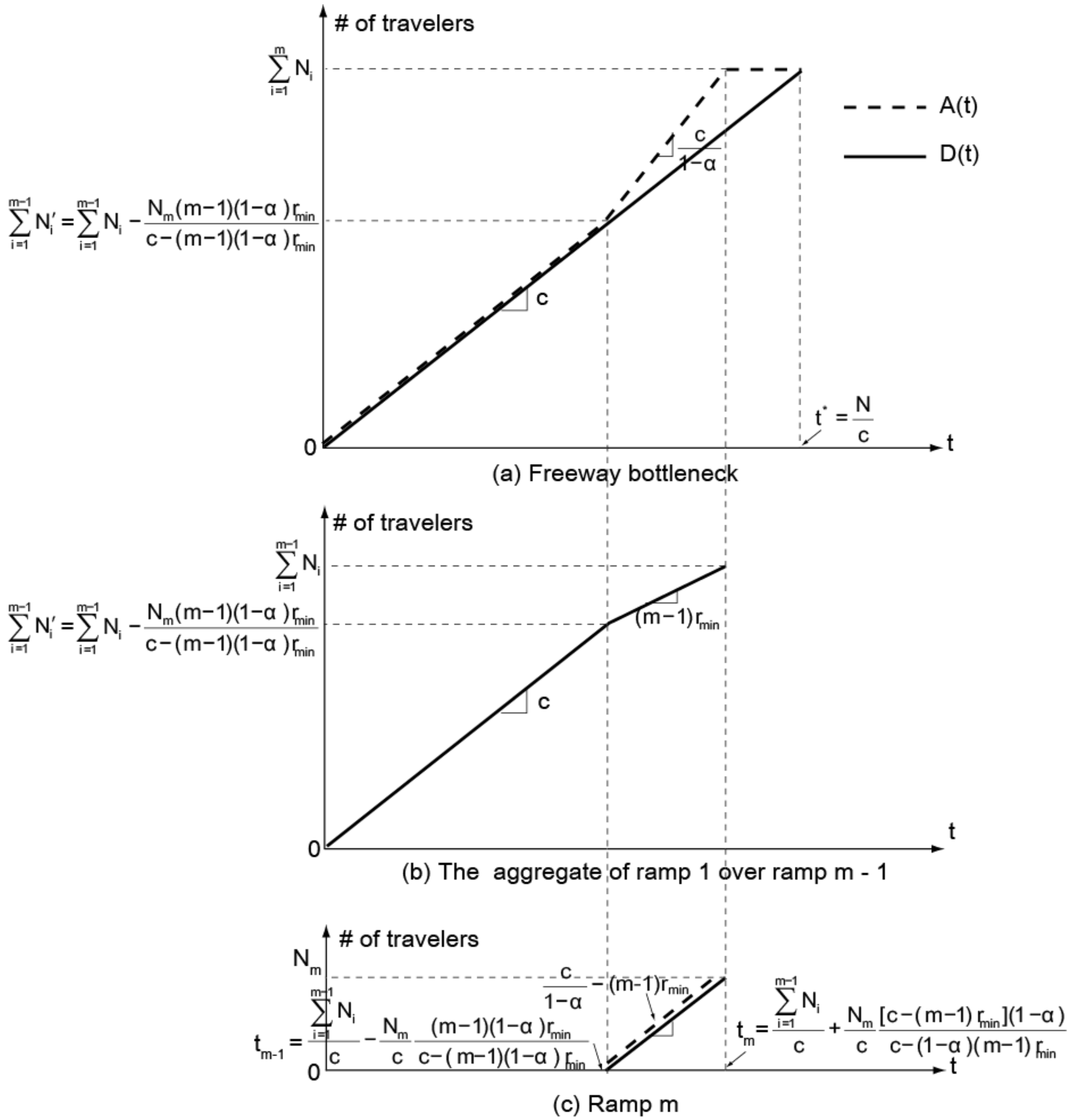


FIGURE 10 The cumulative curves: step I

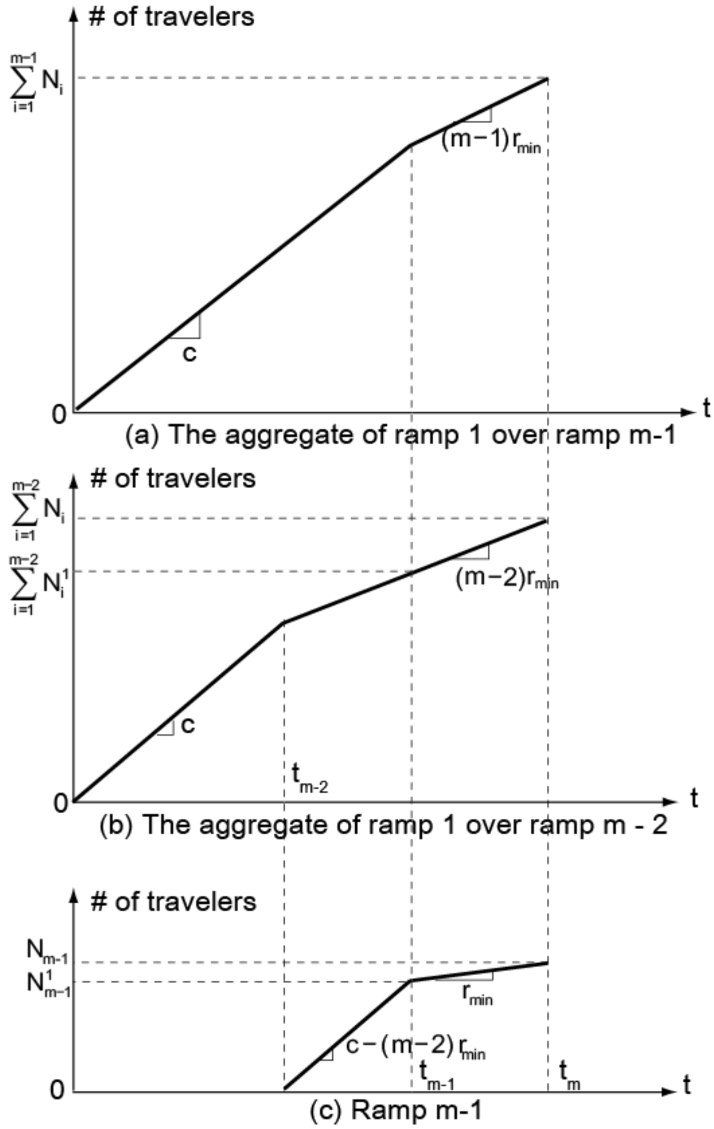


FIGURE 11 The cumulative curves: step II

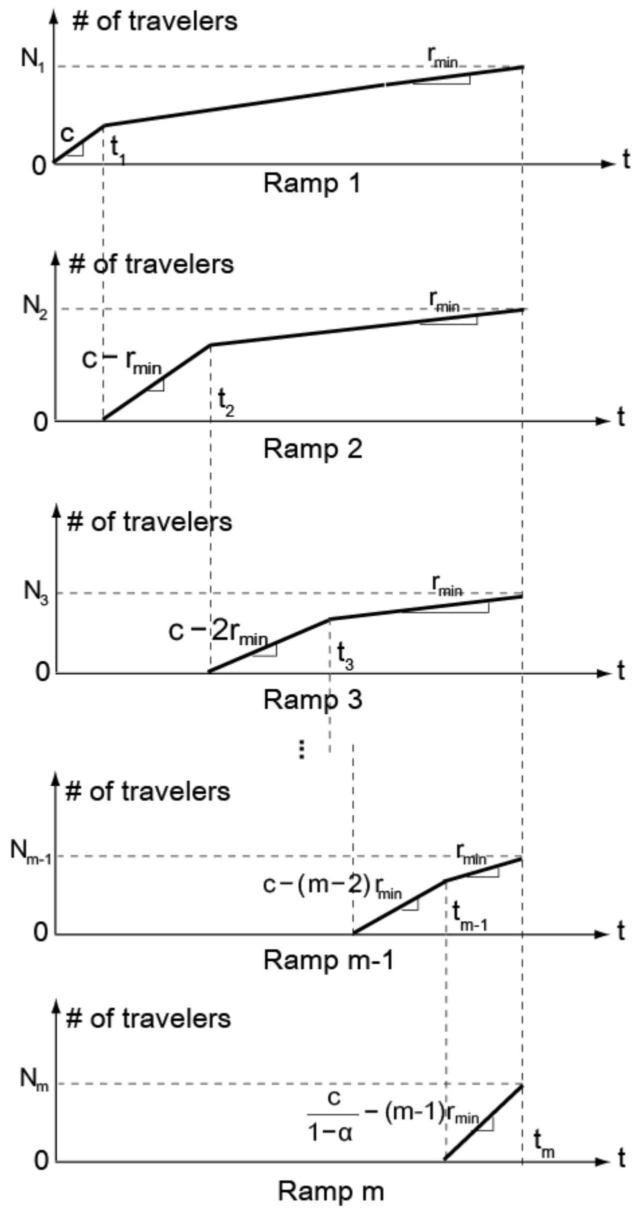


FIGURE 12 The departure curves on each ramp

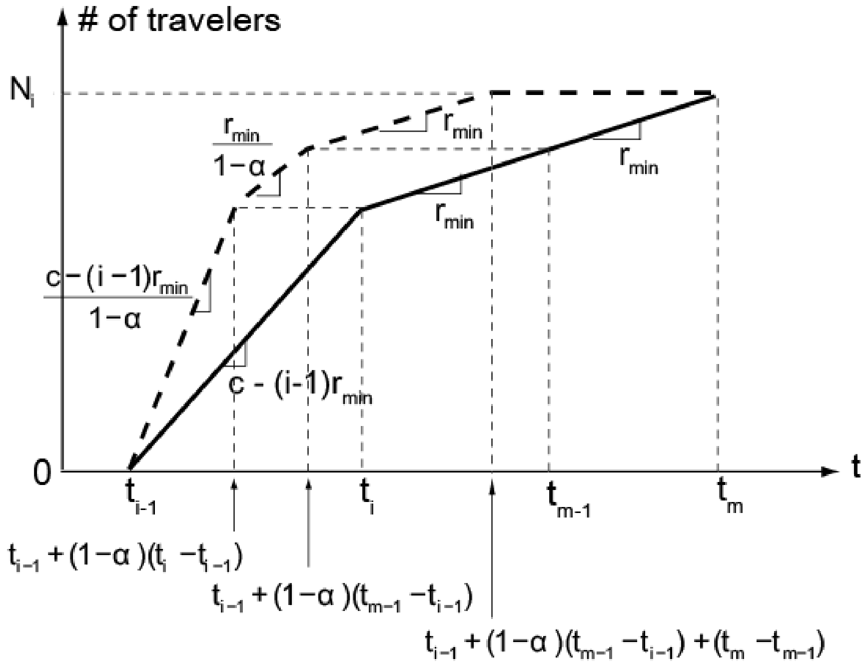


FIGURE 13 The departure and arrival curves on ramp i