12-17-2015

Constrained optimization for bottleneck peak hour pricing

Da Xu

University of Windsor

CONSTRAINED OPTIMIZATION FOR BOTTLENECK PEAK HOUR PRICING

By

Da Xu

A Thesis
Submitted to the Faculty of Graduate Studies
through Mechanical Engineering in Partial Fulfillment
of the Requirements
for the Degree of Master of Applied Science at the
University of Windsor

Windsor, Ontario, Canada

2015

CONSTRAINED OPTIMIZATION FOR BOTTLENECK PEAK HOUR PRICING

by

Da Xu

APPROVED BY:

____
Dr. L. Oriet
Mechanical Engineering

____
Dr. L. Zhong
Odette School of Business

____
Dr. X. Guo, Advisor
Odette School of Business

____
Dr. G. Zhang, Advisor
Mechanical Engineering

November 27, 2015

DECLARATION OF ORIGINALITY

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

We study the morning commute problem with a peak period flat toll, where the toll has a maximum acceptable toll level and a maximum acceptable length of tolling period. Under such a constrained optimization setup, we investigate the system cost minimization problem. A tolling scheme is determined by the toll starting time, the toll ending time, and the toll level. The toll starting time and ending time are set before and after the common work start time, respectively. We find out that, under the toll window length constraint only, a balanced toll window design is always optimal, where "balanced" means that the part of the toll window before the work start time and the part after have equal monetary value. Under both the toll level and the toll window length constraints, the balanced design is optimal if feasible; otherwise the toll should start later with the same toll window length.

ACKNOWLEDGEMENTS

To grandma, for your everlasting love and bless to me.

To my supervisor, Dr. Xiaolei Guo, for guiding me and inspiring me through the completion of this thesis. Thank you for spending so much time talking with me for my life. You are always my respected mentor.

To my supervisor, Dr. Guoqing Zhang, for providing me precious comments to improve this thesis and helping me obtain this degree. I cherish every time we talk in your office. Those moments were happy and unforgettable.

To my committee members, Dr. Leo Oriet and Dr. Ligang Zhong, for your insightful ideas and recommendations to finalize this thesis. Your ideas enlightened me in so many aspects of my research. Thank you for giving me so much inspiration for my future work.

To my beloved parents, for your patients and great love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. Introduction

In recent decades, the problems stemming from high vehicle ownership and heavy road usage has become much starker. These problems include road congestion, pavement damage, air pollution, traffic accidents and limited parking places. As a result, road pricing has been widely implemented all over the world. The world's first congestion tolling scheme was introduced in Singapore's core central business district (CBD) in 1975 as the Singapore Area Licensing Scheme (ALS). The roads leading to the CBD are tolled. If a driver wants to enter the CBD, she needs to purchase a special paper license which is sold at post offices, gas stations or convenience stores, on a monthly or daily basis. The toll gate at the entrance of the CBD are gantries where police officers are visually checking the license and recording any violations. The ALS was upgraded to 100% free-flowing Electronic Road Pricing (ERP) system in September 1998. Sensors are installed on the gantries to communicate with an In-vehicle Unit (IU) to implement the charging. The IU is a device to insert a cash card to pay the toll. When a car drives under a gantry, the sensors on the gantry will work with the IU to deduct the money in the cash card automatically. Each registered car intending to enter the CBD is enforced to install an IU by law. Actually, before Singapore's implementation of ERP, Hong Kong experimented the ERP system during 1983 to 1985. The results demonstrated the technical feasibility of this tolling system, but it was aborted due to the public opposition. In Europe, Norway implemented a cordon tolling scheme in the city of Bergen (1986), Oslo (1990) and Trondheim (1991). The Oslo toll ring is a classic cordon pricing scheme with 19 toll stations circling the centre of Oslo. People driving into the city need to pay a fee when they pass the cordon line. The toll stations support electronic payment without reducing vehicle's speed. The cordon tolling scheme makes every car entering the city centre have to pass a toll station, so the city centre's traffic congestion can be effectively alleviated. The collected toll is intended to improve road network and finance road construction projects. Sweden introduced Stockholm congestion tax that covers Stockholm city centre in August 2007. All the entrances and exits of the centre area have unmanned control points operating with automatic number plate recognition. Vehicles entering this are during the peak hours need to pay a fixed fee. The congestion tax collected from commuters is also used to construct and maintain the toll roads. US first introduced the High-occupancy toll lane (HOT lane) system in 1995 on California's 91 express lanes. In next year, Interstate 15, north of San Diego, also started to implement the HOT tolling scheme. The HOT lane is a traffic lane that is only free to high-occupancy vehicles and designated exempt vehicles. The high occupancy vehicle usually is the vehicle with at least 2 or 3 occupants. Other vehicles intending to use the lane need to pay a toll. If the driver does not like to pay the toll, she can also use the general untolled lane. The toll level is displayed at the entry of the lanes, which is adjusted according to the travel demand to control traffic volume to ensure the minimum traffic speed and service level.

In urban area, during the morning commute peak hours, heavy congestion at road's bottleneck has now become an unneglectable problem for the commuters. It is very common to hear one's colleague complaining how long she has to wait on the road. Since no one would like to come to workplace too early or can afford the penalty of being late, travelers usually depart from home at approximately same time periods. At the road's bottleneck section, due to its capacity limit to handle the travel demand, congestion is inevitable. Motivated by this problem, this thesis proposes a tolling scheme implemented during morning commute peak hours. Our purpose is to alleviate the road congestion caused by the excessive traffic flow through charging a coarse toll at the road's bottleneck area which usually has very limited capacity. We choose to charge a coarse toll because of its easiness to implement. The dynamic toll or the time-varying toll can cause confusion to the commuter, as the commuter may not know when she should depart from home. In this thesis, the bottleneck model is built based on the concept of equilibrium. At equilibrium, no commuter can further reduce her travel cost by altering the arrival time at the bottleneck. By levying a toll on the bottleneck, the equilibrium profile of commuters could have a tremendous change compared with the no-toll scenario. Considering that the toll has a maximum public acceptable level and the tolling period has a maximum public acceptable length, we investigate the problem of system cost minimization and our goal is to find the optimal tolling scheme under these two constraints. Under the optimal tolling scheme, commuters have the minimum total system cost. It could also be understood as the best equilibrium profile of all profiles. Under such a constrained optimization setup, we first solve the equilibrium of the bottleneck model. We find out that, for any toll window, there exists a critical toll level over which capacity waste can happen. Capacity waste is a time period during which, no commuter uses the bottleneck. Then, based on the individual cost, we prove, in respect of total system cost, a tolling scheme without capacity waste is always better than a scheme with capacity waste. We also find out that, under toll window length constraint only, if the unconstrained optimal tolling scheme is infeasible, we should push toll window length to the upper bound, make toll window balanced and charge the corresponding critical toll price. Balanced means the part of the toll window before the work start time and the part after has equal monetary value. Under both toll level and toll window length constraints, if the unconstrained optimal tolling scheme is infeasible, whenever possible, a balanced toll window and its corresponding critical toll price can solve the problem; if the balanced design gives a critical toll price exceeding the upper bound of toll level, we can push the toll window rightward to make the tolling time start later, or namely make the toll window unbalanced, and charge the corresponding critical toll price of the moved toll window.

The remainder of this thesis is organized as follows. In chapter two, we do a literature review to show the previous researches of the road bottleneck model. In chapter three, we review the equilibrium of untolled bottleneck model. Chapter four gives a complete picture of tolled bottleneck model, where we investigated the equilibrium profiles of different tolling schemes. In chapter five, we solve the unconstrained system cost optimization problem based on the individual cost. In chapter six, we solve the constrained system cost optimization problem given both toll level and toll window length constraints. In chapter seven, we use numerical examples to demonstrate the proposed optimal tolling schemes under different constraint setups. Concluding remarks are offered in chapter eight.

2. Literature Review

To model the morning commute problem, Vickery first introduced the road bottleneck model in 1969 (Vickery 1969). Hendrickson and Kocur (1981) reviewed the no-toll equilibrium of bottleneck model with and without no lateness assumption. They also investigated the distribution of work start times and pointed out that the condition of equilibrium is that the arrival rate of commuters must be constant. As bottleneck model's pioneers, Arnott et al. (1988) extend Vickery's model by assuming different work starting times. They investigated users' no-toll equilibrium profile and showed that, for two groups of commuters, the queue at the bottleneck can be single peaked, double peaked or the rush hour can be separate based on the work starting times' difference. Arnott et al. (1990) pointed out by levying a flat toll during the commuting period, travelers' total system cost can be reduced and under the optimal tolling scheme, there should be no queue at the toll window's endpoints. Arnott et al. (1993) extended Vickery's model further by assuming elastic demand and found the optimal road capacity under various pricing regimes. Arnott et al. (1994) examined the welfare effects of the optimal time-varying toll. In their model, the commuters are divided into several groups, each group with its own unique VOT but shares same relative cost of late to early arrival. Under the time-varying toll, queue is completely eliminated but such a tolling scheme depends on each group's VOT and travel demand. In real world it is very inconvenient to implement and can also be quite confusing to commuters, besides their model does not consider continuous VOT distribution either. Further effort was made to reduce commuters' queuing delay at the bottleneck, such as Laih (1994) proposed a multi-step tolling scheme where different flat tolls are levied on different time periods during the peak hour. The flat toll scheme is easier to implement but Laih's model is limited by the assumption that the flat toll does not change commuter's travel cost (compared with no-toll equilibrium). Although under the optimal time-varying toll, the toll revenue equals the saved queuing cost, it is usually not true under a flat toll. Lindsey (2004) reviewed previous bottleneck models under assumption of multiple user classes and proved the existence and uniqueness of user equilibrium of bottleneck model. Xiao et al. (2011) extended Arnott's model (1990) by providing details of how the queuing profile changes with respect to toll level under heterogeneous VOT assumption. They formulate a non-linear optimization problem to solve the equilibrium and find out the optimal tolling scheme. Under their optimal tolling scheme, no queue exists at toll window's endpoints either. This is mainly due to the proportional assumption of user's VOT. Xiao et al. (2013) extend Arnott's model (1994) by assuming continuous VOT distribution, where social optimum is also achieved by a dynamic tolling scheme. Under his model, the toll level only depends on each commuters' VOT and does not require dividing travelers into groups, which can save some work but such a dynamic tolling scheme also suffers from inconvenience of implementing.

Based on our literature review, none of the existing works studied the constrained optimization problem of a tolled bottleneck. For public acceptable issue, we consider that the toll has a maximum acceptable toll level and a maximum acceptable length of tolling period, both exogenously given, so a constrained system optimization problem can be proposed. In this thesis, we still use flat toll for our tolling scheme as it is easy to be implemented in real world. We will solve the constrained system cost optimization problem given both toll level and toll window length constraints and establish the properties of the optimal tolling schemes.

3. Equilibrium of Untolled Single Bottleneck

During the morning commute peak hours, at some busy roads, we can easily observe travelers stuck by heavy congestion. This is usually due to most commuters have roughly same work starting time but the capacity of the road cannot satisfy such high travel demand. In order to model this phenomenon, bottleneck model is introduced. In this section, we will briefly review the equilibrium of untolled single bottleneck during the morning commute period for users with heterogeneous VOT. t^* is assumed to be commuters' preferred arrival time at work. If a commuter arrives at work before t^*, she will be incurred a schedule early delay cost β. If she arrives at work later than t^* she will be incurred a schedule late delay cost γ. The queuing delay cost is denoted by α. We assume $\beta = \alpha\eta$ and $\gamma = \alpha\mu$, where η and μ are constants $\left(0 < \eta < 1 < \mu\right)$. The x th user's VOT $\alpha\left(x\right)$ is assumed to be a monotonically decreasing function with respect to x. The total number of commuters is assumed to be N. The bottleneck's capacity is s. The arrival rate is denoted by λ. When commuter's arrival rate is higher than s, a queue will develop at the bottleneck. When commuter's arrival rate is lower than s, the queue at the bottleneck will gradually dissipate. Since η and μ are both constants, the profile of the no-toll equilibrium should be pretty similar with the equilibrium profile under homogeneous VOT assumption. At equilibrium, the arrival rates of commuters having schedule early and late delay can be obtained as both constants, implying the traveler's position in the queue is random. Since we assume commuters value schedule late delay more than schedule early delay, we can see in figure 1 the arrival rate of commuters having schedule early delay is much higher than that of commuters having schedule late delay. The morning commute period starts at t_q and ends at $t_{q'}$.

We can obtain $t^* - t_q = \mu N / (\eta + \mu) s$ and $t_{q'} - t^* = \eta N / (\eta + \mu) s$. Figure 1 shows the untolled single bottleneck equilibrium profile of users having heterogeneous VOT. (See Appendix A for details)

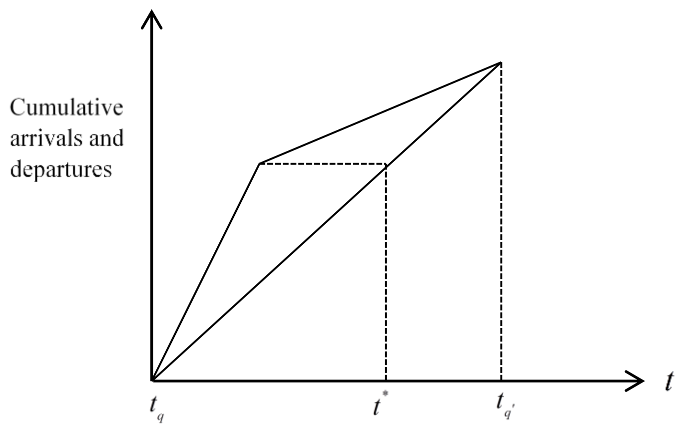

Figure 1: untolled single bottleneck equilibrium profile

4. Equilibrium of Tolled Single Bottleneck

In this section we will talk about the equilibrium of tolled single bottleneck with commuters having heterogeneous VOT. A flat toll p is imposed from t^+ to t^-. Our tolling scheme only assumes $t^+ < t^* < t^-$. At equilibrium no traveler can further reduce her travel cost by adjusting her arrival time at the bottleneck. Since the flat toll has no impact on toll payers' travel time choice, the arrival pattern of toll payers having schedule early delay or late delay should be similar with those under the profile of untolled single bottleneck. As the toll non-payer who arrives before t^+ is incurred a schedule early delay cost, her arrival rate should be similar with that of the toll payer who also has schedule early delay cost. The arrival rate of travelers having schedule early delay can be obtained as $s/1-\eta$. The arrival rate of travelers having schedule late delay can be obtained as $s/1+\mu$.

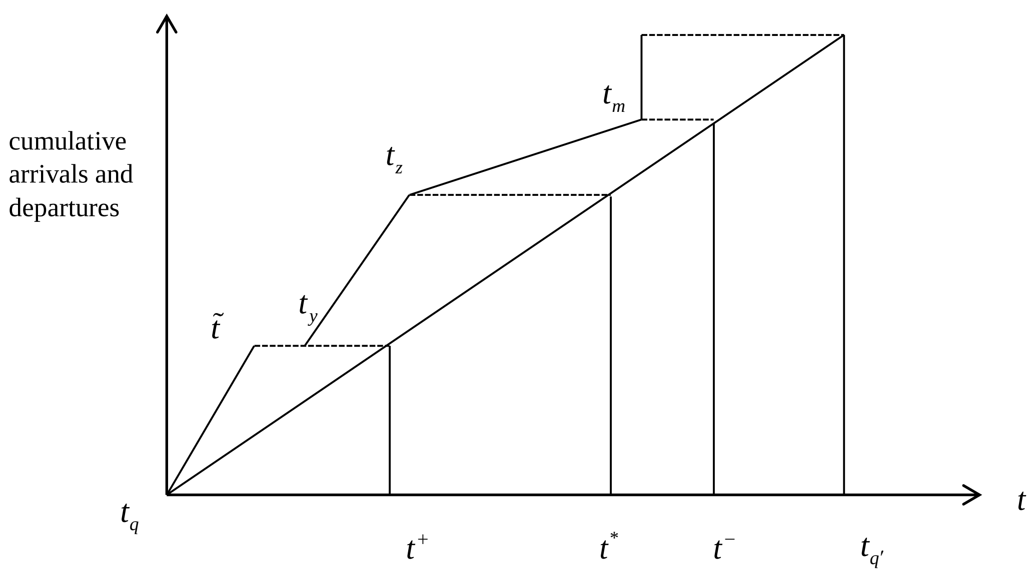

Figure 2: equilibrium profile of tolled bottleneck without capacity waste

When the toll window length is not too long and toll level is not too high, the bottleneck can be fully utilized with no capacity waste. Here capacity waste means a period during which no queue exists at the bottleneck between the arrival time of the first and last commuter. The x th commuter is assumed to be the toll non-payer arriving before t^+, her travel cost can be given by

$$C(x,t)=\alpha(x)\frac{\int_{t_q}^{t}\lambda_1(\omega)d\omega-s(t-t_q)}{s}+\beta(x)\left(t^*-t-\frac{\int_{t_q}^{t}\lambda_1(\omega)d\omega-s(t-t_q)}{s}\right)$$

At equilibrium, we can obtain

$$C(x,t)=\beta(x)(t^*-t_q) \tag{1}$$

As the bottleneck is fully utilized, the first toll payer should come no later than t^+. The y th commuter is assumed to be the toll payer who experiences schedule early delay, her cost can be given by

$$\begin{aligned}C(y,t)&=\alpha(y)\frac{\int_{t_q}^{\tilde{t}}\lambda_1(\omega)d\omega+\int_{t_y}^{t}\lambda_2(\omega)d\omega-s(t-t_q)}{s}\\&+\beta(y)\left(t^*-t-\frac{\int_{t_q}^{\tilde{t}}\lambda_1(\omega)d\omega+\int_{t_y}^{t}\lambda_2(\omega)d\omega-s(t-t_q)}{s}\right)+p\end{aligned}$$

where $\tilde{t}$ is the arrival time of the last toll non-payer arriving before t^+. t_y is the arrival time of first toll payer. At equilibrium, we can obtain

$$C(y,t)=\alpha(y)(t^+-t_y)+\beta(y)(t^*-t^+)+p \tag{2}$$

The last toll payer should arrive no later than t^-. The z th commuter is assumed to be the toll payer who experiences schedule late delay, her cost can be given by

$$C(z,t)=\alpha(z)\frac{\int_{t_q}^{\tilde{t}}\lambda_1(\omega)d\omega+\int_{t_y}^{t_z}\lambda_2(\omega)d\omega+\int_{t_z}^{t}\lambda_3(\omega)d\omega-s(t-t_q)}{s}$$
$$+\gamma(z)\left(t+\frac{\int_{t_q}^{\tilde{t}}\lambda_1(\omega)d\omega+\int_{t_y}^{t_z}\lambda_2(\omega)d\omega+\int_{t_z}^{t}\lambda_3(\omega)d\omega-s(t-t_q)}{s}-t^*\right)+p$$

where traveler who arrives at t_z experiences no schedule early or late delay, since she is cleared just at t^*. At equilibrium, we can obtain

$$C(z,t)=\alpha(z)\left(t^*-t_z\right)+p$$

Mass arrival happens right after the arrival time of the last toll payer. We denote it by t_m. Every commuter in the mass arrival is assumed to experience an average queuing delay and schedule late delay of the total mass. The travel cost of commuter in the mass arrival can be given by

$$C(m,t_m)=\alpha(m)\left(\frac{t^-+t_{q'}}{2}-t_m\right)+\gamma(m)\left(\frac{t^-+t_{q'}}{2}-t^*\right)$$

At equilibrium, the indifferent user is the commuter who can arrive at any time as she is always incurred identical travel cost. For those who have higher VOT than the indifferent user, they will pay the toll to pass the bottleneck. For those who have lower VOT than the indifferent user, they will avoid the toll by coming earlier or later. The toll price can be easily obtained as the queuing cost difference of the indifferent user arriving at $\tilde{t}$ and t_y respectively, thus we have

$$p=\alpha\left[s\left(t^--t^+\right)\right]\left(t_y-\tilde{t}\right)$$

From

$$\frac{s}{1-\eta}\left(\tilde{t}-t_q\right)=s\left(t^+-t_q\right)$$

$$\frac{s}{1-\eta}\left(t_z-t_y\right)=s\left(t^*-t^+\right)$$

$$\frac{s}{1+\mu}(t_m - t_z) = s(t^- - t^*)$$

$$t_{q'} - t_q = \frac{N}{s}$$

we can further obtain

$$t_q = \frac{2p}{\alpha\left[s(t^- - t^+)\right](1+\mu)} + t^- - \frac{N}{s} \tag{3}$$

$$t_y = \frac{2\eta+1+\mu}{2} t_q - (\eta-1)t^+ - \frac{(1+\mu)\left(t^- - \frac{N}{s}\right)}{2}$$

$$t_m = \frac{2\eta+1+\mu}{2} t_q - (\eta+\mu)t^* + \frac{t^- + \frac{N}{s}}{2}(1+\mu)$$

From (3), we can see when toll price is increased, the first commuter will postpone her arrival. Since the first commuter has postponed her arrival, the first and last toll payer will also postpone their arrivals or we can say the equilibrium profile will move rightward. When toll price is increased to a certain level, the first toll payer will arrive exactly at t^+ or the last toll payer will arrive exactly at t^-. At this moment, if we keep increasing the toll level, capacity waste will occur at t^+ or t^-. By setting $t_y = t^+$, we can obtain

$$p_1 = \frac{\alpha\left[s(t^- - t^+)\right]\eta(1+\mu)}{2\eta+1+\mu}\left(t^+ - t^- + \frac{N}{s}\right) \tag{4}$$

If toll window is designed as $\eta(t^* - t^+) > \mu(t^- - t^*)$, $t^- - t^+ < N/s$ and toll level is kept within $0 < p \le p_1$, the bottleneck will be fully utilized. This corresponds to area AOD in figure 2. By setting $t_m = t^-$, we can obtain

$$p_2 = \alpha\left[s\left(t^- - t^+\right)\right]\frac{1+\mu}{2\eta+1+\mu}\left[\eta\frac{N}{s} - (\eta+\mu)\left(t^- - t^*\right)\right] \tag{5}$$

If toll window is designed as $\eta\left(t^* - t^+\right) < \mu\left(t^- - t^*\right)$, $t^- - t^* < \eta N/(\eta+\mu)s$ and toll level is kept within $0 < p \le p_2$, the bottleneck will be fully utilized. This corresponds to area ODF in figure 2.

If toll window is designed balanced as $\eta\left(t^* - t^+\right) = \mu\left(t^- - t^*\right)$, we have $p_1 = p_2$. If toll level is pushed to p_1 or p_2, the first and last toll payer will arrive exactly at t^+ and t^-, which implies no queue exists at the endpoints of the toll window. This corresponds to line OD in figure 2.

When toll level exceeds p_1 or p_2, depending on the design of toll window, capacity waste will start to occur at t^+, t^- or even at both t^+ and t^-. If the toll window satisfies $\eta\left(t^* - t^+\right) > \mu\left(t^- - t^*\right)$, when toll level is pushed above p_1, capacity waste will only occur at t^+. Using the same logic, if the toll window is designed as $\eta\left(t^* - t^+\right) < \mu\left(t^- - t^*\right)$, when toll level exceeds p_2, we will observe capacity waste only at t^-. Of course, for the design of $\eta\left(t^* - t^+\right) = \mu\left(t^- - t^*\right)$, when toll price surpasses p_1 or p_2, capacity waste will occur at both t^+ and t^-.

If capacity waste only exists at t^+, during $\left(t^+, t_y\right)$ no queue exists at the bottleneck. From the standpoint of the indifferent user we can easily obtain the toll price as the travel cost difference of her coming as the first toll non-payer and first toll payer respectively, thus we have

$$p = \alpha(\rho N)\eta\left(t_y - t_q\right)$$

where ρ is the proportion of toll payers. The ρN th commuter is namely the indifferent user. She has the lowest VOT among the toll payers but the highest VOT among the toll non-payers. The toll price is simply the schedule early delay difference of her coming at these two moments. Based on the fundamental equilibrium relations, we can obtain

$$t_m = (\mu+\eta)\left(t^- - t^*\right) + t^- - \frac{\eta\rho N}{s}$$

$$p = \alpha(\rho N)\eta\frac{1+\mu}{1+\mu+2\eta}\left(t^- - t^+ + \frac{N}{s} - 2\frac{\rho N}{s}\right)$$

where

$$\frac{\rho N}{s} = t^- - t_y$$

It is shown the toll price is a function of toll payers' proportion. When toll level is raised up, the indifferent user's VOT will increase. As fewer people can afford the toll, toll payers can gain more time choice freedom. The last toll payer is going to postpone her arrival by arriving closer to t^-. When toll price is raised to a certain level, the last toll payer will arrive exactly at t^-. By setting $t_m = t^-$, we have

$$p_3 = \alpha\left[s\frac{\mu+\eta}{\eta}\left(t^- - t^*\right)\right]\frac{1+\mu}{1+\mu+2\eta}\left[\eta\left(t^* - t^+\right) - (2\mu+\eta)\left(t^- - t^*\right) + \eta\frac{N}{s}\right] \tag{6}$$

If the toll window is designed as $\eta\left(t^* - t^+\right) > \mu\left(t^- - t^*\right)$, $t^- - t^+ < N/s$ and toll price is kept within $p_1 < p \le p_3$, capacity waste only occurs at t^+. This corresponds to area AOD in figure 2.

When toll level exceeds p_3, capacity waste will also occur at t^-. If we have capacity waste at t^-, since there is no queue, the mass arrival should happen exactly at t^-. If we denote $\tilde{t}_z$ as the last toll payer's arrival time, from the standpoint of the indifferent user, we should have

$$\eta\left(t^* - t_y\right) = \mu\left(\tilde{t}_z - t^*\right)$$

This implies the first toll payer's schedule early delay equals the last toll payer's schedule late delay. The toll price can be obtained as

$$p = \alpha\left[s\left(\tilde{t}_z - t_y\right)\right]\frac{(1+\mu)\eta}{1+\mu+2\eta}\left(\frac{2\eta+2\mu}{\eta}t^* - t^+ + \frac{N}{s} + \frac{2\mu}{1+\mu}t^- - \frac{2\mu+2\mu^2+\eta+3\eta\mu}{(1+\mu)\eta}\tilde{t}_z\right)$$

where $\alpha\left[s\left(\tilde{t}_z - t_y\right)\right]$ is the indifferent user's VOT.

It is easy to understand when toll price is pushed high enough, only the zeroth person (commuter with highest VOT) can afford the toll to pass the bottleneck. The rest commuters have to come either before t^+, or after t^-. It is obvious the zeroth commuter should arrive exactly at t^*, since she will not be incurred any schedule delay cost. Such a toll price can be obtained by setting $\tilde{t}_z = t^*$.

By setting $\tilde{t}_z = t^*$, we can obtain

$$p_4 = \alpha(0)\frac{(1+\mu)\eta}{1+\mu+2\eta}\left(\frac{N}{s} + \frac{2\mu}{1+\mu}t^- - t^+ - \frac{\mu-1}{1+\mu}t^*\right) \tag{7}$$

By setting $\tilde{t}_z = t^-$, we can obtain

$$p = \alpha\left[s\frac{\mu+\eta}{\eta}\left(t^- - t^*\right)\right]\frac{1+\mu}{1+\mu+2\eta}\left[\eta\left(t^* - t^+\right) - (2\mu+\eta)\left(t^- - t^*\right) + \eta\frac{N}{s}\right]$$

If the toll window is designed as $\eta\left(t^* - t^+\right) > \mu\left(t^- - t^*\right)$, $t^- - t^+ < N/s$ and toll level is kept within $p_3 < p \le p_4$, capacity waste occurs at both t^+ and t^-. (Figure 2 area AOD)

When toll price is greater than p_4, no commuter can afford the toll, all of them arrive either before t^+ or join the mass arrival at t^-. (Figure 2 area AOD)

Now let us talk about capacity waste only existing at t^-. If the toll window design satisfies $\eta\left(t^* - t^+\right) < \mu\left(t^- - t^*\right)$, when toll level exceeds p_2, capacity waste will only occur at t^-. Since there is no queue between $\tilde{t}_z$ and t^-, mass arrival will happen exactly at t^-. The toll price can be obtained as the indifferent user's travel cost difference of her coming at t_q and t_y respectively, thus we have

$$p = \alpha(\rho N)\eta\left(t^+ - t_q\right) - \alpha(\rho N)\left(t^+ - t_y\right)$$

The toll price can be understood as the travel cost difference of indifferent user arriving as the first toll non-payer and first toll payer respectively. Based on the fundamental equilibrium relations, we can obtain

$$t_y = \eta\left(t^* - t^+\right) - \mu\left(\tilde{t}_z - t^*\right) + t^+$$

$$p = \alpha(\rho N)\frac{2\eta\mu}{2\eta + 1 + \mu}\left[t^- + \frac{\eta + \mu}{2\eta\mu}(1+\mu)t^* + \frac{1+\mu}{2\mu}\frac{N}{s} - \frac{\eta + 3\eta\mu + \mu + \mu^2}{2\eta\mu}\left(t^+ + \frac{\rho N}{s}\right)\right]$$

where

$$\frac{\rho N}{s} = \tilde{t}_z - t^+$$

We can see the toll price is a function of toll payers' proportion. When toll price is increased, the indifferent user's VOT will correspondingly increase. As fewer people can afford the toll, the toll payers can gain more time choice freedom. The first toll payer will postpone her arrival by arriving closer to t^+. Finally when toll price is increased to a certain level, the first toll payer will arrive exactly at t^+ and at this moment if we keep raising the toll level, capacity waste will also occur at t^+. By setting $t_y = t^+$, we could obtain

$$p_5 = \alpha\left[s\frac{\eta+\mu}{\mu}\left(t^*-t^+\right)\right]\left[\frac{2\eta\mu}{2\eta+1+\mu}\left(t^- - t^*\right)-\frac{\eta^2+\eta\mu+\eta\mu^2+3\eta^2\mu}{\left(2\eta+1+\mu\right)\mu}\left(t^*-t^+\right)+\frac{\eta+\eta\mu}{2\eta+1+\mu}\frac{N}{s}\right] \quad (8)$$

If toll window is designed as $\eta\left(t^*-t^+\right)<\mu\left(t^- - t^*\right)$, $t^- - t^* < \eta N/\left(\eta+\mu\right)s$ and toll level is kept within $p_2 < p \le p_5$, capacity waste only occurs at t^-. This corresponds to area ODF in figure 2.

When toll level exceeds p_5, capacity waste will also occur at t^+. Using the same logic, we can see if toll level is kept within $p_5 < p \le p_4$, capacity waste occurs at both t^+ and t^-. (Figure 2 area ODF)

When toll price is higher than p_4, no commuters will use the toll window. All of them will arrive either before t^+ or after t^-. (Figure 2 area ODF)

For the balanced toll window design, we have $p_1 = p_2 = p_3 = p_5$, so when toll level exceeds p_1 but under p_4, capacity waste occurs at both t^+ and t^-. When toll level exceeds p_4, all commuters will arrive either before t^+ or after t^-. (Figure 2 line OD)

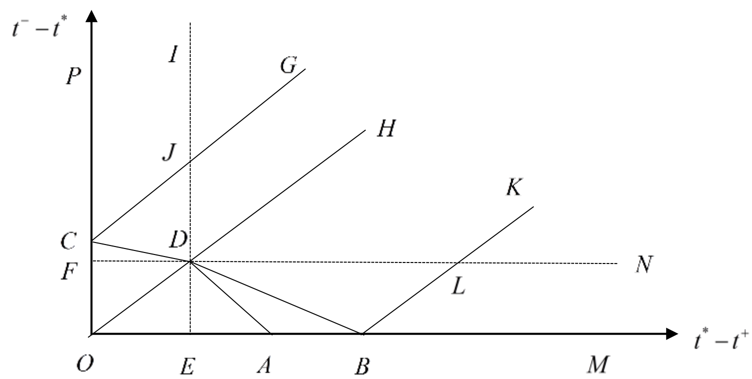

Figure 3: complete tolled single bottleneck equilibrium profile

The following equations are the lines in figure 3.

$$OH: \eta\left(t^* - t^+\right) = \mu\left(t^- - t^*\right),\ \ CG: \mu\left(t^- - t^*\right) - \eta\left(t^* - t^+\right) = \eta\frac{N}{s},\quad D: \left(\frac{\mu}{\eta+\mu}\frac{N}{s}, \frac{\eta}{\eta+\mu}\frac{N}{s}\right)$$

$$BK: \eta\left(t^* - t^+\right) - \mu\left(t^- - t^*\right) = \frac{1+\mu}{2}\frac{N}{s},\ \ AD: \left(t^* - t^+\right) + \left(t^- - t^*\right) = \frac{N}{s}$$

$$BD: \frac{\mu+\eta+\mu^2-\mu\eta}{\eta}\left(t^- - t^*\right) + 2\eta\left(t^* - t^+\right) = \left(1+\mu\right)\frac{N}{s},\ \ CD: \frac{\eta}{\mu}\left(t^* - t^+\right) + \frac{\mu}{\eta}\left(t^- - t^*\right) = \frac{N}{s}$$

The equilibrium profile is not only restricted to what we have discussed above. In the following, we will give a full picture of commuter's equilibrium patterns. (See Appendix B for details) We let

$$p_6 = \alpha\left[\frac{\left(1+\mu\right)N}{1+\mu-2\eta} - \frac{2\eta s}{1+\mu-2\eta}\left(t^- - t^+\right)\right]\frac{\left(1+\mu\right)\eta}{1+\mu-2\eta}\left(t^- - t^+ - \frac{N}{s}\right)$$

$$p_7 = \alpha\left[s\frac{\mu+\eta}{\mu}\left(t^* - t^+\right)\right]\left[\eta\frac{N}{s} - \eta\frac{\left(\eta+\mu\right)}{\mu}\left(t^* - t^+\right)\right]$$

$$p_8=\alpha\left[s\left(t^*-t^+\right)-s\frac{\mu}{\eta}\left(t^- -t^*\right)+N\right]\left[\frac{\mu(\eta+2\mu)}{\eta+\mu}\left(t^- -t^*\right)-\frac{\eta\mu}{\eta+\mu}\left(t^*-t^+ +\frac{N}{s}\right)\right]$$

$$p_9=\alpha(N)\eta\left(t^*-t^+ -\frac{\mu N}{(\eta+\mu)s}\right)$$

$$p_{10}=\alpha(0)\eta\left(t^*-t^+ +\frac{N}{s}\right)$$

$$p_{11}=\alpha\left[s\left(t^*-t^+\right)-s\frac{\mu}{\eta}\left(t^- -t^*\right)+N\right]\left[\mu\left(t^- -t^*\right)+\frac{\mu^2}{\eta}\left(t^- -t^*\right)-\mu\frac{N}{s}\right]$$

$$p_{12}=\alpha\left[s\frac{\eta+\mu}{\eta}\left(t^- -t^*\right)\right]\frac{1+\mu}{2}\left[\frac{N}{s}-\frac{\eta+\mu}{\eta}\left(t^- -t^*\right)\right]$$

$$\begin{aligned}p_{13}=&\alpha\left[N+\frac{2\mu s}{1+\mu}\left(t^- -t^*\right)-\frac{2\eta s}{1+\mu}\left(t^*-t^+\right)\right]\\&\times\left[\frac{\eta(1+\mu)(\mu+\eta)+2\eta^2\mu}{(1+\mu)(\mu+\eta)}\left(t^*-t^+\right)-\frac{2\eta\mu^2}{(1+\mu)(\mu+\eta)}\left(t^- -t^*\right)-\frac{\eta\mu}{\mu+\eta}\frac{N}{s}\right]\end{aligned}$$

$$p_{14}=\alpha(0)\left[\frac{1+\mu}{2}\frac{N}{s}+\mu\left(t^- -t^*\right)\right]$$

$$p_{15}=\alpha(N)\mu\left(t^- -t^* -\frac{\eta N}{(\eta+\mu)s}\right)$$

In area ADB: when $0<p\le p_6$, no commuter arrives before t^+, no capacity waste occurs at t^-, mass arrival occurs before or at t^-; when $p_6<p\le p_3$, commuter arrives before t^+, capacity waste only occurs at t^+, mass arrival occurs before or at t^-; when $p_3<p\le p_4$, commuter arrives before t^+, capacity waste occurs at both t^+ and t^-, mass arrival occurs at t^-.

In area CDJ : when $0 < p \le p_7$, bottleneck can be fully utilized but has no mass arrival; when $p_7 < p < p_8$, capacity waste only occurs at t^+ and there is no mass arrival; when $p_8 < p \le p_4$, mass arrival recurs at t^-, capacity waste occurs at both t^+ and t^-.

In area $GJDH$: when $p_9 < p \le p_8$, capacity waste only occurs at t^+ and there is no mass arrival; when $p_8 < p \le p_4$, mass arrival recurs at t^-, capacity waste occurs at both t^+ and t^-.

In area $PCJI$: when $0 < p \le p_7$, bottleneck can be fully utilized but has no mass arrival; when $p_7 < p \le p_{10}$, capacity waste only occurs at t^+ and there is no mass arrival.

In area IJG : when $p_9 < p \le p_{10}$, capacity waste only occurs at t^+ and there is no mass arrival.

In area CFD : when $0 < p \le p_{11}$, bottleneck can be fully utilized but has no mass arrival; when $p_{11} < p \le p_5$, capacity waste only occurs at t^- and mass arrival occurs at t^-; when $p_5 < p < p_4$, capacity waste occurs at both t^+ and t^-, mass arrival occurs at t^-.

In area LDB : when $0 < p \le p_{12}$, no commuter arrives before t^+, no capacity waste occurs at t^- and mass arrival occurs before or at t^-; when $p_{12} < p \le p_{13}$, no commuter arrives before t^+, capacity waste occurs at t^- and mass arrival occurs at t^-; when $p_{13} < p \le p_4$, commuter arrives before t^+, mass arrival occurs at t^-, capacity waste occurs at both t^+ and t^-.

In area $NLBM$: when $0 < p \le p_{12}$, no commuter arrives before t^+ , no capacity waste occurs at t^- and mass arrival occurs before or at t^- ; when $p_{12} < p \le p_{14}$, no commuter arrives before t^+ , capacity waste occurs at t^- and mass arrival occurs at t^- .

In area $HDLK$: when $p_{15} < p \le p_{13}$, no commuter arrives before t^+ , capacity waste occurs at t^- and mass arrival occurs at t^- ; when $p_{13} < p \le p_4$, commuter arrives before t^+ , mass arrival occurs at t^- , capacity waste occurs at both t^+ and t^- .

In area KLN : when $p_{15} < p \le p_{14}$, no commuter arrives before t^+ , capacity waste occurs at t^- and mass arrival occurs at t^- .

In area $GCOBK$: when $p > p_4$, no commuter uses the toll window, all of them arrive either before t^+ or join the mass arrival at t^- .

In area KBM : when $p > p_{14}$, no commuter uses the toll window, all of them join the mass arrival at t^- .

In area PCG : when $p > p_{10}$, no commuter uses the toll window, all of them come before t^+ .

In area HDN : when $0 < p < p_{15}$, every commuter pays the toll to pass the bottleneck.

In area IDH : when $0 < p < p_9$, every commuter pays the toll to pass the bottleneck.

5. Unconstrained Optimization Problem

In previous section, we have solved the equilibrium of tolled single bottleneck. In this section, we will talk about the unconstrained optimization problem. Unconstrained means there is no constraint on toll price level or toll window length. The goal is to minimize commuters' total queuing delay and total schedule delay. The toll revenue collected from toll payers can be regarded as tax paid to the government, so minimizing the toll revenue is not our concern. Before our discussion, a lemma is introduced:

Lemma 1. For any tolling scheme with capacity waste, by shortening toll window length and reducing toll price, there exists a tolling scheme with no capacity waste and incurring less total cost. (See Appendix C for proof)

The proof of this lemma is complicated but we can understand it in an easy way. The capacity waste happens at t^+ when toll price exceeds p_1 or at t^- when toll price exceeds p_2. If we still want to retain same amount of toll payers, the only way is to stretch the toll window longer, so the original toll payers with relatively lower VOT would still stay within the toll window, because coming earlier for them to avoid the toll would incur a higher schedule early cost. If we reverse this process, for a toll window with capacity waste, we could shorten the toll window length to the clearing period of toll payers (from the first toll payer's clearing time point to last toll payer's clearing time point) and reduce the toll price to a certain level so the same amount of toll payers would still use the tolled bottleneck. As the amount of toll payers does not change, the total queuing and schedule delay of toll payers will not change either. But by shortening the toll window length, the toll non-payers at least can incur less schedule early cost, so the total cost of all commuters will decrease.

Substituting (3) into (1) and (2) gives us

$$C(x,t)=\beta(x)\frac{N}{s}-\beta(x)\left(t^{-}-t^{*}\right)-\beta(x)\frac{2p}{\alpha(\rho N)(1+\mu)} \tag{9}$$

$$C(y,t)=\beta(y)\frac{N}{s}-\beta(y)\left(t^{-}-t^{*}\right)-\alpha(y)\frac{(2\eta+1+\mu)p}{\alpha(\rho N)(1+\mu)}+p \tag{10}$$

(9) and (10) show that, for a fixed toll window, the higher the toll price is charged, the lower a commuter's cost will be, so for any toll window design, we need to push toll level to p_1 or p_2 to achieve the minimal cost.

If the toll window design satisfies $\mu\left(t^{-}-t^{*}\right)<\eta\left(t^{*}-t^{+}\right)$, by setting toll price to p_1, we can obtain

$$C(x,t)=\beta(x)\frac{N}{s}-\beta(x)\left(t^{-}-t^{*}\right)-\beta(x)\frac{2\eta}{2\eta+1+\mu}\left(t^{+}-t^{-}+\frac{N}{s}\right) \tag{11}$$

$$C(y,t)=\alpha(y)\eta\left(t^{*}-t^{+}\right)+p \tag{12}$$

If the toll window design satisfies $\mu\left(t^{-}-t^{*}\right)>\eta\left(t^{*}-t^{+}\right)$, by setting toll price to p_2, we can obtain

$$C(x,t)=\beta(x)\frac{(1+\mu)N}{(2\eta+1+\mu)s}+\beta(x)\left(t^{-}-t^{*}\right)\frac{\mu-1}{2\eta+1+\mu} \tag{13}$$

$$C(y,t)=\alpha(y)\mu\left(t^{-}-t^{*}\right)+p \tag{14}$$

If the toll window design satisfies $\mu\left(t^{-}-t^{*}\right)=\eta\left(t^{*}-t^{+}\right)$, by setting toll price to p_1 or p_2, the commuters' travel cost can be expressed by either form of the unbalanced design.

In this section, our goal is to achieve the unstrained optimality, so we do not need to worry about the toll level or toll window length. As shown above, there are three different designs of toll window, so our concern is which one will incur the lowest cost. The logic is to pick up a toll window, design it in three different ways, by comparing the individual user's travel cost, we can find the best design.

Let us first compare the design of $\mu\left(t^{-}-t^{*}\right)<\eta\left(t^{*}-t^{+}\right)$ and $\mu\left(t^{-\prime}-t^{*}\right)=\eta\left(t^{*}-t^{+\prime}\right)$, where $t^{-}-t^{+}=t^{-\prime}-t^{+\prime}=\rho N/s$. We readily have

$$t^{-}-t^{*}<\frac{\eta}{\eta+\mu}\frac{\rho N}{s},\ \ t^{-\prime}-t^{*}=\frac{\eta}{\eta+\mu}\frac{\rho N}{s}$$

$$t^{*}-t^{+}>\frac{\mu}{\eta+\mu}\frac{\rho N}{s},\ \ t^{*}-t^{+\prime}=\frac{\mu}{\eta+\mu}\frac{\rho N}{s}$$

Based on (11) and (12), we can see the balanced design is better.

Now let us compare the design of $\mu\left(t^{-}-t^{*}\right)>\eta\left(t^{*}-t^{+}\right)$ and $\mu\left(t^{-\prime}-t^{*}\right)=\eta\left(t^{*}-t^{+\prime}\right)$, where $t^{-}-t^{+}=t^{-\prime}-t^{+\prime}=\rho N/s$. We readily have

$$t^{-}-t^{*}>\frac{\eta}{\eta+\mu}\frac{\rho N}{s},\ \ t^{-\prime}-t^{*}=\frac{\eta}{\eta+\mu}\frac{\rho N}{s}$$

Based on (13) and (14), we can see the balanced design is still better, so for the unconstrained optimization we need to design the toll window balanced and push toll price to p_1 or p_2. The next question is how long the toll window should be. To determine the optimal toll window length, we need to solve

$$\min TC(\rho)=\int_{\rho N}^{N}\frac{\beta(x)(1+\mu)N}{(2\eta+1+\mu)s}+\frac{\beta(x)\eta\rho N}{(\eta+\mu)s}\frac{\mu-1}{2\eta+1+\mu}dx+\int_{0}^{\rho N}\alpha(y)\mu\frac{\eta}{\eta+\mu}\frac{\rho N}{s}dy \qquad (15)$$

(15) is the total queuing delay and schedule delay of all commuters under the balanced toll window design. It is easy to see that total cost is a function of toll payers' proportion ρ. For ease of exposition, we define the following two terms

$$A(x)=\int\alpha(x)dx,\ B(x)=\int\beta(x)dx$$

It is obvious $B(x)=\eta A(x)$. Since VOT is greater than zero, both $A(x)$ and $B(x)$ should be increasing functions with respect to x.

From (15), we can obtain

$$\frac{dTC}{d\rho}=\frac{B(N)\eta N(\mu-1)}{(\mu+\eta)(2\eta+1+\mu)s}-\frac{\beta(\rho N)(1+\mu)N^2}{(2\eta+1+\mu)s}-\frac{\eta N(\mu-1)(\beta(\rho N)\rho N+B(\rho N))}{(\mu+\eta)(2\eta+1+\mu)s}$$
$$+\frac{\mu\eta}{\eta+\mu}\frac{N}{s}\left(\alpha(\rho N)N\rho+A(\rho N)\right)-A(0)\frac{\mu\eta}{\eta+\mu}\frac{N}{s}$$

When $\rho=0$, we can acquire

$$\frac{dTC}{d\rho}=\left(B(N)-B(0)\right)\frac{\eta N(\mu-1)}{(\mu+\eta)(2\eta+1+\mu)s}-\beta(0)\frac{(1+\mu)N^2}{(2\eta+1+\mu)s}$$

Based on Lagrange mean value theorem, we can obtain

$$B(N)-B(0)=\beta(\varepsilon)N\quad(0<\varepsilon<N)$$

obviously $\beta(\varepsilon)<\beta(0)$ and

$$\frac{\eta(\mu-1)}{(\mu+\eta)(2\eta+1+\mu)}<\frac{1+\mu}{2\eta+1+\mu}$$

so we readily have

$$\frac{dTC}{d\rho}(\rho=0)<0$$

When $\rho=1$, we can acquire

$$\frac{dTC}{d\rho}=\alpha(N)\frac{\mu\eta}{\eta+\mu}\frac{N^2}{s}-\beta(N)\frac{(1+\mu)N^2}{(2\eta+1+\mu)s}-\frac{\beta(N)\eta(\mu-1)N^2}{(\mu+\eta)(2\eta+1+\mu)s}+\left(A(N)-A(0)\right)\frac{\mu\eta}{\eta+\mu}\frac{N}{s}$$

since it holds

$$\alpha(N)\frac{\mu\eta}{\eta+\mu}\frac{N^2}{s}-\beta(N)\frac{(1+\mu)N^2}{(2\eta+1+\mu)s}-\frac{\beta(N)\eta(\mu-1)N^2}{(\mu+\eta)(2\eta+1+\mu)s}=0$$

we readily have

$$\frac{dTC}{d\rho}(\rho=1)>0$$

The second order derivative with respect to ρ can be given by

$$\begin{aligned}\frac{d^2TC}{d\rho^2}=&-\beta'(\rho N)\frac{(1+\mu)N^3}{(2\eta+1+\mu)s}-\frac{\eta N(\mu-1)}{(\mu+\eta)(2\eta+1+\mu)s}\left(\beta'(\rho N)\rho N^2+2\beta(\rho N)N\right)\\&+\frac{\mu\eta}{\eta+\mu}\frac{N}{s}\left(\alpha'(\rho N)\rho N^2+2\alpha(\rho N)N\right)\end{aligned}$$

We can further obtain

$$\frac{d^2TC}{d\rho^2}=-\beta'(\rho N)\frac{(1+\mu)N^3}{(2\eta+1+\mu)s}+\beta'(\rho N)\rho\frac{(1+\mu)N^3}{(2\eta+1+\mu)s}+\frac{2\beta(\rho N)(1+\mu)N^2}{(2\eta+1+\mu)s}$$

which readily gives us

$$\frac{d^2TC}{d\rho^2}>0$$

As $TC(\rho)$ is a continuous function, these characteristics guarantee $TC(\rho)$ is a convex function within interval $[0,1]$ so there must be a global minimizer ρ^*. It is apparent that within $\left[0,\rho^*\right]$,

$TC(\rho)$ is a monotonically decreasing function and within $[\rho^*, 1]$, $TC(\rho)$ is a monotonically increasing function. The optimal solution of unconstrained optimization is given by

$$\rho = \rho^*, \quad p = \frac{\alpha(\rho^* N)\eta(1+\mu)}{2\eta + 1 + \mu}\left(\frac{N}{s} - \frac{\rho^* N}{s}\right)$$

6. Constrained Optimization Problem

In previous section, we have discussed how to set the tolling scheme to achieve optimality with no constraint imposed on toll window length or toll price. But in real world, from political standpoint, the toll window length cannot be set too long and the toll price level cannot be charged too high, so now our job becomes how to minimize users' time costs under these constraints. The toll window length constraint can be given by $\rho N/s \le l \ \left(l < N/s\right)$, where parameter l is a pre-determined toll window length limit. The toll price constraint can be given by $p \le \bar{p}$, where parameter $\bar{p}$ is a pre-determined toll price limit.

The first problem we consider is how to achieve optimality with only time constraint on toll window length but no price constraint on toll level. Such a concern is reasonable since the morning commute flow only lasts for one or two hours, if we charge the whole commuting period, the congestion tolling will be pointless. In previous section, we have proved for any toll window, in order to achieve optimality, we need to make it balanced ($\eta\left(t^* - t^+\right) = \mu\left(t^- - t^*\right)$) and push toll level to p_1. Since here we do not have any constraint on toll level either, we also need to make the toll window balanced and push toll level to p_1. The optimization problem can be given by

$$\min TC\left(\rho\right) = \int_{\rho N}^{N} \beta\left(x\right)\frac{\left(1+\mu\right)N}{\left(2\eta+1+\mu\right)s} + \beta\left(x\right)\frac{\eta\rho N}{\left(\eta+\mu\right)s}\frac{\mu-1}{2\eta+1+\mu}dx + \int_{0}^{\rho N} \alpha\left(y\right)\mu\frac{\eta}{\eta+\mu}\frac{\rho N}{s}dy$$

Subject to

$$\frac{\rho N}{s} \le l$$

Since within $\left[0,\rho^*\right]$, $TC(\rho)$ is a monotonically decreasing function, we can obtain the following result: if $\rho^* N/s \le l$, ρ^* is the solution; if $\rho^* N/s > l$, in order to minimize the commuters' total cost, we need to push toll window length to l, namely, l is the solution.

The second problem we consider is how to achieve optimality with both time constraint on toll window length and price constraint on toll level. Differing from the scenario with only time constraint, the balanced toll window design may not be feasible. We need to compare the unconstrained optimal solution with our constraints l and $\overline{p}$ to determine the tolling scheme. Totally four scenarios can be developed in this problem.

6.1. Scenario one: $\rho^* N/s < l$, $p^* < \overline{p}$

We can see this scenario is the easiest scenario, as the unconstrained optimal solution is covered by both the constraints, so ρ^* and p^* are the solution.

6.2. Scenario two: $\rho^* N/s > l$, $p^* < \overline{p}$

In this scenario, although p^* is still within the range of our constraint, toll window length has exceeded the limit. The first step of the optimization is to design the toll window balanced and push toll window length to l, based on (4), if it holds

$$\frac{\alpha(ls)\eta(1+\mu)}{2\eta+1+\mu}\left(\frac{N}{s}-l\right) \le \overline{p}$$

since $TC(\rho)$ is a decreasing function on $\left[0,\rho^*\right]$, the solutions are l and

$$\frac{\alpha(ls)\eta(1+\mu)}{2\eta+1+\mu}\left(\frac{N}{s}-l\right)$$

If it holds

$$\frac{\alpha(ls)\eta(1+\mu)}{2\eta+1+\mu}\left(\frac{N}{s}-l\right)>\bar{p} \tag{16}$$

we can see under the design of $\mu\left(t^{-}-t^{*}\right)\leq\eta\left(t^{*}-t^{+}\right)$, based on (9) and (10), we need to charge the toll price as higher as possible, so the toll price should be taken as $\bar{p}$. For same toll window length, if designed balanced, we have

$$t^{-}-t^{*}=\frac{\eta}{\eta+\mu}\frac{\rho N}{s}$$

If designed as $\mu\left(t^{-}-t^{*}\right)<\eta\left(t^{*}-t^{+}\right)$, we have

$$t^{-}-t^{*}<\frac{\eta}{\eta+\mu}\frac{\rho N}{s}$$

Based on (9) and (10), when toll price is same, the balanced design will incur a lower individual cost, so the design of $\mu\left(t^{-}-t^{*}\right)<\eta\left(t^{*}-t^{+}\right)$ is ruled out. We only need to consider the balanced design. Now we need to solve:

$$\begin{aligned}\min TC(\rho)=&\int_{0}^{\rho N}\beta(y)\frac{N}{s}-\beta(y)\frac{\eta\rho N}{(\eta+\mu)s}-\alpha(y)\frac{(2\eta+1+\mu)\,\bar{p}}{\alpha(\rho N)(1+\mu)}dy\\&+\int_{\rho N}^{N}\beta(x)\frac{N}{s}-\beta(x)\frac{\eta\rho N}{(\eta+\mu)s}-\beta(x)\frac{2\bar{p}}{\alpha(\rho N)(1+\mu)}dx\end{aligned}$$

Subject to

$$\frac{\rho N}{s}\leq l$$

The total cost under the balanced design with toll price taken as $\overline{p}$ is still a function of ρ. We can further obtain $TC(\rho)$ as

$$TC(\rho)=\left[B(N)-B(0)\right]\left(\frac{N}{s}-\frac{\eta\rho N}{(\eta+\mu)s}\right)-\frac{2\overline{p}\left[B(N)-B(\rho N)\right]}{\alpha(\rho N)(1+\mu)}$$
$$-\frac{(2\eta+1+\mu)\,\overline{p}\left[A(\rho N)-A(0)\right]}{\alpha(\rho N)(1+\mu)}$$

We can obtain

$$\frac{dTC}{d\rho}=-\left[B(N)-B(0)\right]\frac{\eta N}{(\eta+\mu)s}+$$
$$\overline{p}\,\frac{2\left[B(N)-B(\rho N)\right]\alpha'(\rho N)N-(1+\mu)\alpha^{2}(\rho N)N+(2\eta+1+\mu)\left[A(\rho N)-A(0)\right]\alpha'(\rho N)N}{\alpha^{2}(\rho N)(1+\mu)}$$

Since $\alpha'(\rho N)<0$, we readily have

$$\frac{dTC}{d\rho}<0$$

This tells us $TC(\rho)$ is a monotonically decreasing function. To minimize $TC(\rho)$, we need to push toll window length to l.

We have considered the design of $\mu\left(t^{-}-t^{*}\right)\leq\eta\left(t^{*}-t^{+}\right)$, now let us talk about the design of $\mu\left(t^{-}-t^{*}\right)>\eta\left(t^{*}-t^{+}\right)$. Based on (9) and (10), we can see for a toll window $\left(t^{+},t^{-}\right)$, if p_2 is lower than $\overline{p}$, then the toll price should be taken as p_2; if p_2 is greater than $\overline{p}$, the toll price should be taken as $\overline{p}$. Let us first consider the scenario of p_2 lower than $\overline{p}$. We need to minimize:

$$\min TC\left(\rho,t^{-}-t^{*}\right)=\int_{0}^{\rho N}\alpha(y)\mu\left(t^{-}-t^{*}\right)dy+\int_{\rho N}^{N}\beta(x)\frac{(1+\mu)N}{(2\eta+1+\mu)s}+\beta(x)\left(t^{-}-t^{*}\right)\frac{\mu-1}{2\eta+1+\mu}dx$$

Subject to

$$\frac{\alpha(\rho N)(1+\mu)}{2\eta+1+\mu}\left[\eta\frac{N}{s}-(\eta+\mu)\left(t^{-}-t^{*}\right)\right]\leq\overline{p} \tag{17}$$

$$\frac{\rho N}{s}\leq l$$

The total cost can be given as

$$TC\left(\rho,t^{-}-t^{*}\right)=\left[A(\rho N)-A(0)\right]\mu\left(t^{-}-t^{*}\right)+\frac{(1+\mu)N\left[B(N)-B(\rho N)\right]}{(2\eta+1+\mu)s}$$
$$+\frac{(\mu-1)\left(t^{-}-t^{*}\right)\left[B(N)-B(\rho N)\right]}{2\eta+1+\mu}$$

The total cost can be treated as a function of $\left(t^{-}-t^{*}\right)$ and ρ. We can acquire

$$\frac{\partial TC\left(\rho,t^{-}-t^{*}\right)}{\partial\rho}=\alpha(\rho N)N\mu\left(t^{-}-t^{*}\right)-\beta(\rho N)N\frac{(1+\mu)N}{(2\eta+1+\mu)s}-\beta(\rho N)N\left(t^{-}-t^{*}\right)\frac{\mu-1}{2\eta+1+\mu}$$
$$=\alpha(\rho N)N\frac{(1+\mu)(\eta+\mu)}{(2\eta+1+\mu)}\left(t^{-}-t^{*}\right)-\alpha(\rho N)N\frac{\eta(1+\mu)N}{(2\eta+1+\mu)s}$$

which readily gives us

$$\frac{\partial TC\left(\rho,t^{-}-t^{*}\right)}{\partial\rho}<0$$

This tells us that under the design of $\mu\left(t^{-}-t^{*}\right)>\eta\left(t^{*}-t^{+}\right)$, when t^{-} is fixed, in order to lower the total cost, we need to stretch t^{+} as left as possible.

We can also acquire

$$\frac{\partial TC\left(\rho,t^{-}-t^{*}\right)}{\partial\left(t^{-}-t^{*}\right)}=\left[A(\rho N)-A(0)\right]\mu+\left[B(N)-B(\rho N)\right]\frac{\mu-1}{2\eta+1+\mu}$$

It is easy to see

$$\frac{\partial TC\left(\rho,t^{-}-t^{*}\right)}{\partial\left(t^{-}-t^{*}\right)}>0$$

This tells us for fixed toll window length, we need to move the toll window as left as possible.

Based on (16), we can see when toll window length equals l, if we move toll window rightward, $t^- - t^*$ will increase and $t^* - t^+$ will decrease. When the toll window is moved rightward to a certain point, there must exist such $\left(t^+, t^-\right)$ which can satisfy

$$\frac{\alpha\left(ls\right)\left(1+\mu\right)}{2\eta+1+\mu}\left[\eta\frac{N}{s}-\left(\eta+\mu\right)\left(t^- - t^*\right)\right]=\bar{p} \tag{18}$$

At this time, if we continue to move toll window rightward, $t^- - t^*$ will become even longer. Since it holds

$$\frac{\partial TC\left(\rho, t^- - t^*\right)}{\partial\left(t^- - t^*\right)} > 0,$$

the total cost will increase.

If toll window length decreases (less than l), in order to make

$$\frac{\alpha\left(\rho N\right)\left(1+\mu\right)}{2\eta+1+\mu}\left[\eta\frac{N}{s}-\left(\eta+\mu\right)\left(t^- - t^*\right)\right]\le \bar{p},$$

based on (18), since $\alpha\left(x\right)$ is a decreasing function, we need to increase $t^- - t^*$ and decrease $t^* - t^+$. Since we have

$$\frac{\partial TC\left(\rho, t^- - t^*\right)}{\partial \rho} < 0,$$

we can see the total cost under such a toll window is higher than a toll window with same length of $t^- - t^*$ but total length of l.

These tell us to minimize the total cost under design of $\mu\left(t^{-}-t^{*}\right)>\eta\left(t^{*}-t^{+}\right)$ and constraint (17), we need to push toll window length to l and charge a toll of $\overline{p}$ or namely we need to find a toll window that satisfies (18).

Now let us talk about the scenario with p_2 greater than $\overline{p}$. We need to solve

$$\min TC\left(\rho,t^{-}-t^{*}\right)=\int_{\rho N}^{N}\beta\left(x\right)\frac{N}{s}-\beta\left(x\right)\left(t^{-}-t^{*}\right)-\beta\left(x\right)\frac{2\overline{p}}{\alpha\left(\rho N\right)\left(1+\mu\right)}dx$$
$$+\int_{0}^{\rho N}\beta\left(y\right)\frac{N}{s}-\beta\left(y\right)\left(t^{-}-t^{*}\right)-\alpha\left(y\right)\frac{\overline{p}\left(2\eta+1+\mu\right)}{\alpha\left(\rho N\right)\left(1+\mu\right)}dy$$

Subject to

$$\frac{\alpha\left(\rho N\right)\left(1+\mu\right)}{2\eta+1+\mu}\left[\eta\frac{N}{s}-\left(\eta+\mu\right)\left(t^{-}-t^{*}\right)\right]>\overline{p} \tag{19}$$

$$\frac{\rho N}{s}\leq l$$

Total cost can be given by

$$TC\left(\rho,t^{-}-t^{*}\right)=\left[B\left(N\right)-B\left(0\right)\right]\left[\frac{N}{s}-\left(t^{-}-t^{*}\right)\right]-\frac{2\overline{p}\left[B\left(N\right)-B\left(\rho N\right)\right]}{\alpha\left(\rho N\right)\left(1+\mu\right)}$$
$$-\frac{\overline{p}\left(2\eta+1+\mu\right)\left[A\left(\rho N\right)-A\left(0\right)\right]}{\alpha\left(\rho N\right)\left(1+\mu\right)}$$

We can acquire

$$\frac{\partial TC\left(\rho,t^{-}-t^{*}\right)}{\partial\rho}<0 \tag{20}$$

This tells us with fixed t^{-}, in order to minimize total cost, we need to stretch t^{+} as left as possible. It is also easy to obtain

$$\frac{\partial TC\left(\rho,t^{-}-t^{*}\right)}{\partial\left(t^{-}-t^{*}\right)}<0 \tag{21}$$

This tells us with fixed toll window length, to minimize total cost, we need to move the toll window as right as possible.

As $\rho N/s=\left(t^{-}-t^{*}\right)+\left(t^{*}-t^{+}\right)$, total cost can also be treated as a function of $t^{-}-t^{*}$ and $t^{*}-t^{+}$, or namely, $TC\left(t^{-}-t^{*},t^{*}-t^{+}\right)$, we can obtain

$$\frac{\partial TC\left(t^{*}-t^{+},t^{-}-t^{*}\right)}{\partial\left(t^{-}-t^{*}\right)}=-\left[B(N)-B(0)\right]+$$
$$\bar{p}\frac{2\left[B(N)-B(\rho N)\right]\alpha'(\rho N)s-(1+\mu)\alpha^{2}(\rho N)s+(2\eta+1+\mu)\left[A(\rho N)-A(0)\right]\alpha'(\rho N)s}{\alpha^{2}(\rho N)(1+\mu)}$$

which readily gives us

$$\frac{\partial TC\left(t^{*}-t^{+},t^{-}-t^{*}\right)}{\partial\left(t^{-}-t^{*}\right)}<0 \tag{22}$$

This tells us with fixed t^{+}, in order to minimize total cost, we need to stretch t^{-} as right as possible.

As long as (18) holds (has solution), for a toll window whose critical toll price is bigger than $\bar{p}$ ($p_2>\bar{p}$), we can always find another toll window whose critical toll price equals $\bar{p}$ ($p_2=\bar{p}$) by stretching t^{-} rightward and/or moving the toll window rightward. Based on (9) and (10), it is shown that with same toll level and same length of toll window, a longer $t^{-}-t^{*}$ will incur a lower individual system cost. This tells us under constraint $p\le\bar{p}$, the system cost of a toll window whose critical toll price equals $\bar{p}$ is lower than that of a toll window with same length but whose critical toll price is greater than $\bar{p}$ (we can move the toll window with $p_2>\bar{p}$ rightward to obtain a toll window with $p_2 = \bar{p}$). If by only moving the toll rightward, we cannot find a toll window whose critical toll price equals $\bar{p}$, we can first stretch t^- rightward to certain point and then move it rightward, so finally we can obtain a toll window whose critical toll price equals $\bar{p}$. Based on (22), (9) and (10), we can see that, if (18) holds, for a toll window whose critical toll price is bigger than $\bar{p}$, there must exits a toll window whose critical toll price equals $\bar{p}$ with less system cost. This tells us a toll window that satisfies (18) is better than any toll window whose critical toll price is bigger than $\bar{p}$.

If (18) has no solution or it holds that

$$\frac{\alpha(ls)(1+\mu)}{2\eta+1+\mu}\left[\eta\frac{N}{s}-(\eta+\mu)l\right] > \bar{p},$$

based on (20), (21) and (22), to minimize the total system cost, we need to push toll window length to l, take the toll price as $\bar{p}$ and move the toll window as right as possible, or namely, the optimal toll window (t^+, t^-) is $(t^*, t^* + l)$.

Now in scenario two under (16) we have two sub-optimal solutions: the first one is to push toll window length to l, make it balanced and charge a toll of $\bar{p}$; the second one is to find a toll window that satisfies (18) or a toll window of $(t^*, t^* + l)$. We need to compare them to find the best one. Based on (9) and (10), it is obvious that with same toll level and same length of toll window, a longer $t^- - t^*$ will incur a lower individual system cost. For the balanced design, we have

$$t^- - t^* = \frac{\eta}{\eta+\mu}l$$

For the unbalanced design, we have

$$t^{-} - t^{*} > \frac{\eta}{\eta + \mu} l$$

So it is straightforward that the unbalanced design has a lower system cost than the balanced design. Now our conclusion is that if (18) has solution, the optimal solution is a toll window satisfies (18); if (18) has no solution, the optimal solution is to push the toll window length to l and move it as right as possible or namely, a toll window of $\left(t^{*}, t^{*} + l\right)$. Actually, when (18) has no solution, it implies the political constraints imposed on the tolling scheme are too strong. It implicates the toll window length limit is too short compared to the toll level limit or the toll level limit is too low compared to the toll window length limit.

6.3. Scenario three: $\rho^{*} N/s < l, p^{*} > \bar{p}$

In this scenario, the unconstrained optimal toll window length is within our limit, but the toll price has exceeded the limit. In order to find the optimal solution under such constraints, the first step we should do is to solve

$$\frac{\alpha(\rho N)\eta(1+\mu)}{2\eta + 1 + \mu}\left[\frac{N}{s} - \left(t^{-} - t^{+}\right)\right] = \bar{p} \tag{23}$$

we denote the solution as $\bar{\rho}$

If $\bar{\rho} N/s \le l$, for $\rho \ge \bar{\rho}$, under the balanced design, $\rho = \bar{\rho}$ is the optimal solution, because (15) is an increasing function with respect to ρ on interval $\left[\rho^{*}, 1\right]$. Under the balanced design, for $\rho < \bar{\rho}$, we have

$$\frac{\alpha(\rho N)\eta(1+\mu)}{2\eta+1+\mu}\left[\frac{N}{s}-\left(t^- - t^+\right)\right] > \bar{p}\ ,$$

so the toll price should be charged by $\bar{p}$. With the toll price being $\bar{p}$, under balanced design, total cost is a decreasing function with respect to ρ, so the minimal cost will still be achieved by taking $\rho = \bar{\rho}$. Now we need to think about the unbalanced design $\mu\left(t^- - t^*\right) > \eta\left(t^* - t^+\right)$. We already proved in the section of unconstrained optimization that for the same toll window length, the balanced design will incur the minimal cost. So for the toll window with length $\bar{\rho}N/s$, the balanced design with toll price being p_1 is always better than the unbalanced design with toll price being p_2. And for $\rho = \bar{\rho}$, we can always find a dummy toll level bound $\bar{p}' < \bar{p}$, then find a toll window that can satisfy

$$\frac{\alpha(\bar{\rho}N)(1+\mu)}{2\eta+1+\mu}\left[\eta\frac{N}{s}-(\eta+\mu)\left(t^- - t^*\right)\right] = \bar{p}'$$

We have proved in scenario two that such a toll window is the optimal solution for problem with constraints $\rho \le \bar{\rho}$ and $p \le \bar{p}'$.

Based on these discussion, we can see the optimal toll window length should be $\bar{\rho}N/s$, and toll price should be $\bar{p}$, or namely a toll window that satisfies (23).

If $\bar{\rho}N/s > l$, thus for every $\rho N/s < l$, we have

$$\frac{\alpha(\rho N)\eta(1+\mu)}{2\eta+1+\mu}\left[\frac{N}{s}-\left(t^- - t^+\right)\right] > \bar{p}$$

Based on the discussion of scenario two, the optimal solution is a toll window that should satisfy equation (18). Of course, if (18) has no solution, we can minimize total cost as much as possible by making toll window unbalanced, pushing toll window length to l and moving it as right as possible or namely a toll window of $\left(t^*, t^* + l\right)$.

6.4. Scenario four: $\rho^* N/s > l,\ p^* > \bar{p}$

In the last scenario, the unconstrained optimal toll window length and toll level have both exceeded the limits, it is obvious that

$$\frac{\alpha(ls)\eta(1+\mu)}{2\eta+1+\mu}\left(\frac{N}{s}-l\right) > \bar{p},$$

based on the discussion of scenario two, the optimal solution is a toll window that should satisfy equation (18). If (18) has no solution, we can minimize total cost as much as possible by making toll window unbalanced, pushing toll window length to l and moving it as right as possible or namely a toll window of $\left(t^*, t^* + l\right)$.

7. Numerical Examples

We use a numerical example to demonstrate what tolling scheme should be adopted under both toll window length and toll level constraints. We use the VOT setup adopted in Xiao et al.'s paper (2011). The travel demand is 100. Capacity of the bottleneck is 50. $\eta = 0.609$ and $\mu = 2.377$. The preferred arrival time at work place is at 0 o'clock (the time points only serve as reference points). The difference is that the VOT function in our model is a monotonically decreasing function. Following is the VOT function setup:

$$\alpha(x) = 0.128(100 - x)$$

$$\beta(x) = 0.078(100 - x)$$

$$\gamma(x) = 0.3042(100 - x)$$

Table 1 shows the unconstrained optimal tolling scheme

Table 1: unconstrained optimal tolling scheme	
proportion of toll payers	39.91%
toll window length	0.798
tolling starting time	-0.635
tolling ending time	0.163
optimal toll level	4.14
peak hour stating time	-1.5182
peak hour ending time	0.488

We can see that under the unconstrained optimal tolling scheme the peak hour starts at -1.5182 and ends at 0.488. We need to start charging the toll at -0.635 and stop charging at 0.163. The optimal toll level is 4.14. There will be almost 40 toll payers.

If there is only toll window length constraint, e.g. $l = 0.75$, the optimal tolling scheme is shown in table 2.

Table 2: constrained optimal tolling scheme with $l = 0.75$	
proportion of toll payers	37.50%
toll window length	0.75
tolling starting time	-0.597
tolling ending time	0.153
optimal toll level	4.48
peak hour stating time	-1.515
peak hour ending time	0.485

We can see the unconstrained optimal tolling scheme is no longer feasible, but with constraint only on toll window length, we still adopt the balanced toll window design.

If we have both toll window length and toll level constraint, e.g. $l = 0.75$ and $\overline{p} = 4.3$, the optimal solution is shown in table 3.

Table 3: constrained optimal tolling scheme with $l = 0.75$ and $\overline{p} = 4.3$	
proportion of toll payers	37.50%
toll window length	0.75
tolling starting time	-0.587
tolling ending time	0.163
optimal toll level	4.3
peak hour stating time	-1.5817
peak hour ending time	0.4813

The toll window length constraint in table 3 is same with table 2 and the two schemes have same number of toll payers. But with the toll level constraint of table 3, we need to start charging the toll later.

Table 4, Table 5, Table 6 and Table 7 give another four examples of constrained optimal tolling scheme under both toll window length and toll level constraints.

Table 4: constrained optimal tolling scheme with $l = 0.85$ and $\bar{p} = 4$	
proportion of toll payers	40.90%
toll window length	0.818
tolling starting time	-0.651
tolling ending time	0.167
optimal toll level	4
peak hour stating time	-1.52
peak hour ending time	0.48

Table 5: constrained optimal tolling scheme with $l = 0.75$ and $\bar{p} = 4$	
proportion of toll payers	37.50%
toll window length	0.75
tolling starting time	-0.57
tolling ending time	0.18
optimal toll level	4
peak hour stating time	-1.524
peak hour ending time	0.476

Table 6: constrained optimal tolling scheme with $l = 0.75$ and $\overline{p} = 4.6$	
proportion of toll payers	37.50%
toll window length	0.75
tolling starting time	-0.597
tolling ending time	0.153
optimal toll level	4.48
peak hour stating time	-1.515
peak hour ending time	0.485

Table 7: constrained optimal tolling scheme with $l = 0.81$ and $\overline{p} = 4$	
proportion of toll payers	40.5%
toll window length	0.81
tolling starting time	-0.6414
tolling ending time	0.1686
optimal toll level	4
peak hour stating time	-1.5203
peak hour ending time	0.4797

8. Conclusions

This thesis studied the morning commute problem with a peak period flat toll on a single bottleneck. We first solved the equilibrium of bottleneck model under different tolling schemes. We found out that, for any toll window, there exists a critical toll level over which capacity waste happens. Then, based on the individual cost, we proved that, in respect of total system cost, a tolling scheme without capacity waste is always better than a scheme with capacity waste. We also found out that, under toll window length constraint only, if the unconstrained optimal tolling scheme is infeasible, we should push toll window length to the upper bound, make toll window balanced and charge the corresponding critical toll price. Balanced means the part of the toll window before the work start time and the part after has equal monetary value. Under both toll level and toll window length constraints, if the unconstrained optimal tolling scheme is infeasible, whenever possible, a balanced toll window and its corresponding critical toll price can solve the problem; if the balanced design gives a critical toll price exceeding the upper bound of toll level, we can push the toll window rightward to make the tolling time start later, or namely make the toll window unbalanced, and charge the corresponding critical toll price of the moved toll window.

References

Vickrey W. S. 1969. Congestion theory and transport investment. American Economic Review, 59, 251-261.

Hendrickson, C., G. Kocur 1981. Schedule delay and departure time decisions in a deterministic model. Transportation Science 15(1) 62-77.

Arnott R, Palma AD, Lindsey R 1988. Schedule delay and departure time decisions with heterogeneous commuters. Transportation Research Record 1197:56–67.

Arnott, R., De Palma, A., Lindsey, R., 1990. Economics of a bottleneck. Journal of Urban Economics 27(1), 111-130.

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

Arnott R, Palma AD, Lindsey R 1994. A welfare congestion tolls with heterogeneous commuters. Journal of Transportation Economics and Policy 28:139–161.

Laih CH (1994) Queuing at a bottleneck with single-Step and multistep tolls. Transportation Research Part A 28:197–208.

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

Ramadurai, G., Ukkusuri, S., Zhao, J., Pang, J., 2010. Linear complementarity formulation for single bottleneck model with heterogeneous commuters. Transportation Research Part B 44 (2), 193–214.

Xiao, F., Qian, Z., Zhang, H.M., 2011. The morning commute problem with coarse toll and nonidentical commuters. Networks and Spatial Economics 11(2), 343–369.

Xiao, F., Shen, W., Zhang, H.M., 2012. The morning commute under flat toll and tactical waiting. Transportation Research Part B 46 (10), 1346-1359.

Xiao, F., Qian, Z., Zhang, H.M., 2013. Managing bottleneck congestion with tradable credits. Transportation Research Part B 56, 1-14.

Appendix A. Derivation of Untolled Bottleneck Equilibrium

As the bottleneck is not tolled, commuters can be divided into two groups. One has queuing delay and schedule early delay. The other has queuing delay and schedule late delay. We assume the x th user has the schedule early delay and the y th user has the schedule late delay. Her cost is given by

$$C(x,t) = \alpha(x)\left(\frac{\int_{t_q}^{t}\lambda_1(\omega)d\omega - s\left(t-t_q\right)}{s}\right) + \beta(x)\left(t^* - t - \frac{\int_{t_q}^{t}\lambda_1(\omega)d\omega - s\left(t-t_q\right)}{s}\right)$$

$$\int_{t_q}^{\tilde{t}}\lambda_1(\omega)d\omega = s\left(t^* - t_q\right) \tag{24}$$

Condition (24) implies the x th commuters are cleared from t_q to t^*. By setting $\partial C(x,t)/\partial t = 0$, we can obtain

$$\lambda_1(t) = \frac{s}{1-\eta}$$

Substituting $\lambda_1(t)$ into $C(x,t)$ gives us

$$C(x,t) = \alpha(x)\eta(t^* - t_q)$$

The y th user's cost is given by

$$\begin{aligned} C(y,t) = {} & \alpha(y)\left(\frac{\int_{t_q}^{\tilde{t}}\lambda_1(\omega)d\omega + \int_{\tilde{t}}^{t}\lambda_2(\omega)d\omega - s(t-t_q)}{s}\right) \\ & +\gamma(y)\left(t + \frac{\int_{t_q}^{\tilde{t}}\lambda_1(\omega)d\omega + \int_{\tilde{t}}^{t}\lambda_2(\omega)d\omega - s\left(t-t_q\right)}{s} - t^*\right) \end{aligned}$$

$$\int_{\tilde{t}}^{t_{q'}}\lambda_2(\omega)d\omega = s\left(t_{q'} - t^*\right) \tag{25}$$

Condition (25) implies the y th users are cleared from t^* to $t_{q'}$. By setting $\partial C(y,t)/\partial t = 0$, we can obtain

$$\lambda_2(t) = \frac{s}{1+\mu}$$

Substituting $\lambda_2(t)$ into $C(y,t)$ gives us

$$C(y,t) = \alpha(y)(\tilde{t} - t_q)\frac{\eta+\mu}{1-\eta} - \alpha(y)\mu(t^* - t_q)$$

Substituting $\lambda_1(t)$ and $\lambda_2(t)$ into (24) and (25) can give us

$$s(t^* - t_q) = \frac{s}{1-\eta}(\tilde{t} - t_q) \tag{26}$$

$$s(t_{q'} - t^*) = \frac{s}{1+\mu}(t_{q'} - \tilde{t}) \tag{27}$$

Based on (26), (27) and the fact $s(t_{q'} - t_q) = N$, we can readily obtain

$$t_q = t^* - \frac{\mu N}{(\eta+\mu)s}$$

$$\tilde{t} = t^* - \frac{\eta\mu N}{(\eta+\mu)s}$$

$$t_{q'} = t^* + \frac{\eta N}{(\eta+\mu)s}$$

If the x th commuter chose to come after $\tilde{t}$, he would have a cost of

$$\alpha(x)(\tilde{t} - t_q)\frac{\eta+\mu}{1-\eta} - \alpha(x)\mu(t^* - t_q)$$

Based on (26) and (27), we can obtain

$$\alpha(x)\eta(t^* - t_q) = \alpha(x)(\tilde{t} - t_q)\frac{\eta+\mu}{1-\eta} - \alpha(x)\mu(t^* - t_q)$$

This tells us at equilibrium, a commuter can choose to arrive at any time. No matter when she comes, she will have the same cost.

Appendix B. Derivation of tolled bottleneck equilibrium

Here, we give the derivation of the equilibrium profile of capacity waste at both t^+ and t^- as an example.

Capacity waste happening at both t^+ and t^-:

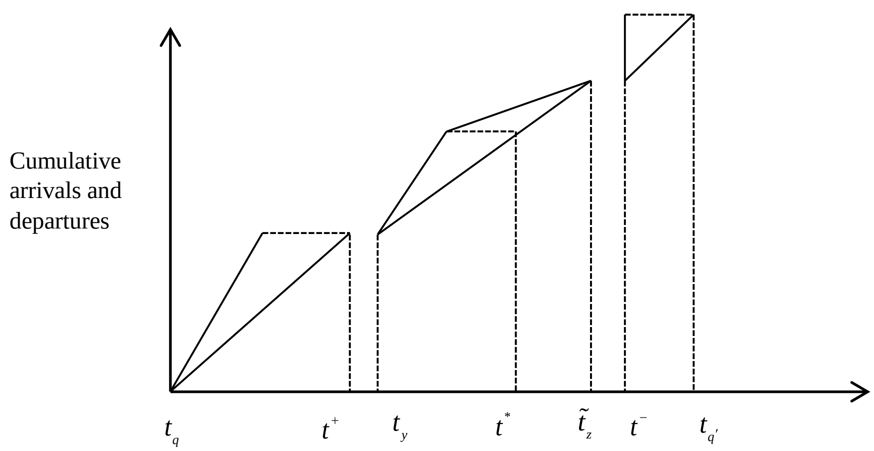

Figure 4: equilibrium profile of capacity waste at both t^+ and t^-

If there is capacity waste at both t^+ and t^-, it is very obvious that the mass arrival happens at t^-. The first toll payer arrives after t^+ and the last toll payer arrives before t^-. Each traveler's individual cost is given as

$$C(x,t)=\alpha(x)\frac{\int_{t_q}^{t}\lambda_1(\omega)d\omega-s(t-t_q)}{s}+\beta(x)\left(t^*-t-\frac{\int_{t_q}^{t}\lambda_1(\omega)d\omega-s(t-t_q)}{s}\right)$$

$$C(y,t)=\alpha(y)\frac{\int_{t_y}^{t}\lambda_2(\omega)d\omega-s(t-t_y)}{s}+\beta(y)\left(t^*-t-\frac{\int_{t_y}^{t}\lambda_2(\omega)d\omega-s(t-t_y)}{s}\right)+p$$

$$C(z,t) = \alpha(z)\frac{\int_{t_y}^{t_z}\lambda_2(\omega)d\omega + \int_{t_z}^{t}\lambda_3(\omega)d\omega - s(t-t_y)}{s}$$
$$+\gamma(z)\left(t + \frac{\int_{t_y}^{t_z}\lambda_2(\omega)d\omega + \int_{t_z}^{t}\lambda_3(\omega)d\omega - s(t-t_y)}{s} - t^*\right) + p$$

$$C(m,t^-) = \alpha(m)\left(\frac{t_{q'} - t^-}{2}\right) + \gamma(m)\left(\frac{t^- + t_{q'}}{2} - t^*\right)$$

By setting $\partial C(x,t)/\partial t = 0$, $\partial C(y,t)/\partial t = 0$ and $\partial C(z,t)/\partial t = 0$, we can obtain

$$\lambda_1(t) = \lambda_2(t) = \frac{s}{1-\eta}$$

$$\lambda_3(t) = \frac{s}{1+\mu}$$

Substituting $\lambda_1(t)$, $\lambda_2(t)$ and $\lambda_3(t)$ into to $C(x,t)$, $C(y,t)$ and $C(z,t)$ gives us

$$C(x,t) = \beta(x)(t^* - t_q)$$

$$C(y,t) = \beta(y)(t^* - t_y) + p$$

$$C(z,t) = \alpha(z)(t^* - t_z) + p$$

At equilibrium, no commuter can further reduce his individual cost by changing arrival time, so we must have

$$\beta(x)(t^* - t_q) < \beta(x)(t^* - t_y) + p$$

$$\beta(y)(t^* - t_y) + p < \beta(y)(t^* - t_q)$$

$$\beta(x)(t^* - t_q) \le \alpha(x)\left(\frac{t_{q'} - t^-}{2}\right) + \gamma(x)\left(\frac{t^- + t_{q'}}{2} - t^*\right)$$

$$\alpha(m)\left(\frac{t_{q'}-t^-}{2}\right)+\gamma(m)\left(\frac{t^-+t_{q'}}{2}-t^*\right)\le\beta(m)\left(t^*-t_q\right)$$

$$\beta(y)\left(t^*-t_y\right)+p\le\alpha(y)\left(t^*-t_z\right)+p$$

$$\alpha(z)\left(t^*-t_z\right)+p\le\beta(z)\left(t^*-t_y\right)+p$$

which readily gives us

$$\alpha(x)<\frac{p}{\eta\left(t_y-t_q\right)} \tag{28}$$

$$\alpha(y)>\frac{p}{\eta\left(t_y-t_q\right)} \tag{29}$$

$$\eta\left(t^*-t_y\right)=t^*-t_z$$

$$\eta\left(t^*-t_q\right)=\left(\frac{t_{q'}-t^-}{2}\right)+\mu\left(\frac{t^-+t_{q'}}{2}-t^*\right) \tag{30}$$

From the fact that

$$\frac{s}{1-\eta}\left(\tilde{t}-t_q\right)=s\left(t^+-t_q\right)$$

$$\frac{s}{1-\eta}\left(t_z-t_y\right)=s\left(t^*-t_y\right)$$

$$\frac{s}{1+\mu}\left(\tilde{t}_z-t_z\right)=s\left(\tilde{t}_z-t^*\right)$$

we can further obtain

$$\tilde{t}=t^+-t^+\eta+t_q\eta \tag{31}$$

$$t_z=t^*-t^*\eta+t_y\eta \tag{32}$$

$$t_z=t^*+t^*\mu-\tilde{t}_z\mu \tag{33}$$

(32) and (33) readily gives us

$$\eta\left(t^* - t_y\right) = \mu\left(\tilde{t}_z - t^*\right)$$

The ρN th person is the indifferent user who can choose to arrive at any time at equilibrium. Based on (28) and (29), we readily have

$$\alpha(\rho N) = \frac{p}{\eta\left(t_y - t_q\right)}$$

this can give us

$$t_q = t_y - \frac{p}{\alpha(\rho N)\eta} \tag{34}$$

The toll payers are cleared from t_y to $\tilde{t}_z$, we can easily obtain

$$\tilde{t}_z - t_y = \frac{\rho N}{s}$$

The length of the peak hour can be given by

$$t_{q'} - t_q = \frac{N}{s} + \left(t_y - t^+\right) + \left(t^- - \tilde{t}_z\right) \tag{35}$$

which readily gives us

$$t_{q'} - t^- = t_q + \frac{N}{s} + t_y - t^+ - \tilde{t}_z \tag{36}$$

$$t_{q'} + t^- = t_q + \frac{N}{s} + t_y - t^+ + 2t^- - \tilde{t}_z \tag{37}$$

Substituting (36) and (37) into (30) can give us

$$(\eta + \mu)t^* - \eta t_q = \frac{t_q + \frac{N}{s} + t_y - t^+ - \tilde{t}_z}{2}(1 + \mu) + \mu t^- \tag{38}$$

Substituting (34) into (38) gives us

$$\eta\left(t^{*}-t_{y}\right)+\mu\left(t^{*}-t_{y}\right)-t_{y}+\frac{p}{\alpha\left(\rho N\right)}\frac{1+\mu+2\eta}{2\eta}=\frac{N}{2s}\left(1+\mu\right)-\frac{\left(t^{+}+\tilde{t}_{z}\right)\left(1+\mu\right)}{2}+\mu t^{-} \tag{39}$$

since $\eta\left(t^{*}-t_{y}\right)=\mu\left(\tilde{t}_{z}-t^{*}\right)$, (39) can further be converted to

$$\left(1+\mu\right)\left(\tilde{t}_{z}-t_{y}\right)+\frac{p}{\alpha\left(\rho N\right)}\frac{1+\mu+2\eta}{\eta}-\frac{N}{s}\left(1+\mu\right)=2\mu\left(t^{-}-\tilde{t}_{z}\right)+\left(1+\mu\right)\left(t_{y}-t^{+}\right) \tag{40}$$

since $\tilde{t}_{z}-t_{y}=\rho N/s$, (40) can give us

$$\frac{p}{\alpha\left(\rho N\right)}\frac{1+\mu+2\eta}{\left(1+\mu\right)\eta}-\left(1-\rho\right)\frac{N}{s}=\frac{2\mu}{\left(1+\mu\right)}\left(t^{-}-\tilde{t}_{z}\right)+\left(t_{y}-t^{+}\right) \tag{41}$$

Based on the fact $\eta\left(t^{*}-t_{y}\right)=\mu\left(\tilde{t}_{z}-t^{*}\right)$ and (41), we can obtain

$$p=\alpha\left[s\left(\tilde{t}_{z}-t_{y}\right)\right]\frac{\left(1+\mu\right)\eta}{1+\mu+2\eta}\left(\frac{2\eta+2\mu}{\eta}t^{*}-t^{+}+\frac{N}{s}+\frac{2\mu}{1+\mu}t^{-}-\frac{2\mu+2\mu^{2}+\eta+3\eta\mu}{\left(1+\mu\right)\eta}\tilde{t}_{z}\right) \tag{42}$$

The condition bottleneck has capacity waste on both sides is given by

$$\tilde{t}<t^{+}<t_{y}$$

$$\tilde{t}_{z}<t^{-}<t_{q'}$$

Since $\eta\left(t^{*}-t_{y}\right)=\mu\left(\tilde{t}_{z}-t^{*}\right)$ and $\tilde{t}_{z}<t^{-}$, we can easily obtain

$$\tilde{t}_{z}<t^{*}+\frac{\eta}{\mu}\left(t^{*}-t^{+}\right)$$

Also from $t^{-}<t_{q'}$ and (35), we can obtain

$$t^{-}<\frac{N}{s}+t_{y}-t^{+}+t^{-}-\tilde{t}_{z}+t_{q}$$

which gives us

$$\tilde{t}_{z}<\frac{N}{s}+t_{y}-t^{+}+t_{q} \tag{43}$$

Based on (34) and (42), (43) can further give us

$$\tilde{t}_z < \frac{\eta}{\mu+\eta}\frac{N}{s} + 2t^* - \frac{\eta}{\mu+\eta}t^+ - \frac{\mu}{\mu+\eta}t^- \tag{44}$$

As $\tilde{t}_z > t^*$, we readily have

$$\frac{\eta}{\mu+\eta}\frac{N}{s} + 2t^* - \frac{\eta}{\mu+\eta}t^+ - \frac{\mu}{\mu+\eta}t^- > t^*$$

which gives us

$$\eta\left(t^* - t^+\right) - \mu\left(t^- - t^*\right) + \eta\frac{N}{s} > 0$$

From $t^+ > \tilde{t}$ and (31) , we can easily obtain

$$t^+ > t^+ - t^+\eta + t_q\eta$$

this gives us $t^+ > t_q$

which readily gives us

$$t^+ > t_y - \frac{p}{\alpha(\rho N)\eta}$$

so we can further obtain

$$\tilde{t}_z < \frac{2\eta^2}{(\mu+\eta)(1+\mu)}t^+ + \frac{1+\mu-2\eta}{1+\mu}t^* + \frac{\eta}{\mu+\eta}\frac{N}{s} + \frac{2\mu\eta}{(\mu+\eta)(1+\mu)}t^- \tag{45}$$

As $\tilde{t}_z > t^*$, we readily have

$$\frac{2\eta^2}{(\mu+\eta)(1+\mu)}t^+ + \frac{1+\mu-2\eta}{1+\mu}t^* + \frac{\eta}{\mu+\eta}\frac{N}{s} + \frac{2\mu\eta}{(\mu+\eta)(1+\mu)}t^- > t^*$$

this gives us

$$\mu\left(t^- - t^*\right) - \eta\left(t^* - t^+\right) + \frac{1+\mu}{2}\frac{N}{s} > 0$$

so finally we need

$$-\frac{1+\mu}{2}\frac{N}{s} < \mu\left(t^- - t^*\right) - \eta\left(t^* - t^+\right) < \eta\frac{N}{s}$$

Now let us talk about the toll price's range under different designs of toll window.

Let the right side of (44) minus right side of (45) gives us

$$\frac{1+\mu+2\eta}{(1+\mu)(\mu+\eta)}\left[\eta\left(t^* - t^+\right) - \mu\left(t^- - t^*\right)\right]$$

When $\eta\left(t^* - t^+\right) > \mu\left(t^- - t^*\right)$**, we need**

$$\tilde{t}_z < t^- \text{ and } \tilde{t}_z < \frac{2\eta^2}{(\mu+\eta)(1+\mu)}t^+ + \frac{1+\mu-2\eta}{1+\mu}t^* + \frac{\eta}{\mu+\eta}\frac{N}{s} + \frac{2\mu\eta}{(\mu+\eta)(1+\mu)}t^-$$

If it holds

$$\frac{\mu+\eta+\mu^2-\mu\eta}{\eta}\left(t^- - t^*\right) + 2\eta\left(t^* - t^+\right) < \frac{N}{s}(1+\mu),$$

we need

$$t^* < \tilde{t}_z < t^-$$

We can also acquire t_y's range as

$$t^* - \frac{\mu}{\eta}\left(t^- - t^*\right) < t_y < t^*$$

From the plotting of line

$$\frac{\mu+\eta+\mu^2-\mu\eta}{\eta}\left(t^- - t^*\right) + 2\eta\left(t^* - t^+\right) - \frac{N}{s}(1+\mu) = 0$$

and line

$$\eta\left(t^* - t^+\right) - \mu\left(t^- - t^*\right) = 0,$$

we can see we further need

$$t^* - t^+ < \frac{1+\mu}{2\eta}\frac{N}{s} \text{ and } t^- - t^* < \frac{\eta}{\mu+\eta}\frac{N}{s}$$

Based on (42), toll price's range can be given by

$$p > \alpha\left[s\frac{\mu+\eta}{\eta}\left(t^- - t^*\right)\right]\frac{1+\mu}{1+\mu+2\eta}\left[\eta\left(t^* - t^+\right) - \left(2\mu+\eta\right)\left(t^- - t^*\right) + \eta\frac{N}{s}\right]$$

and

$$p < \alpha\left(0\right)\frac{\left(1+\mu\right)\eta}{1+\mu+2\eta}\left(\frac{N}{s} + \frac{2\mu}{1+\mu}t^- - t^+ - \frac{\mu-1}{1+\mu}t^*\right)$$

If it holds

$$\frac{\mu+\eta+\mu^2-\mu\eta}{\eta}\left(t^- - t^*\right) + 2\eta\left(t^* - t^+\right) > \frac{N}{s}\left(1+\mu\right),$$

we need

$$t^* < \tilde{t}_z < \frac{2\eta^2}{\left(\mu+\eta\right)\left(1+\mu\right)}t^+ + \frac{1+\mu-2\eta}{1+\mu}t^* + \frac{\eta}{\mu+\eta}\frac{N}{s} + \frac{2\mu\eta}{\left(\mu+\eta\right)\left(1+\mu\right)}t^-$$

From the plotting of the line

$$\frac{\mu+\eta+\mu^2-\mu\eta}{\eta}\left(t^- - t^*\right) + 2\eta\left(t^* - t^+\right) - \frac{N}{s}\left(1+\mu\right) = 0$$

and the line

$$\eta\left(t^* - t^+\right) - \mu\left(t^- - t^*\right) = 0$$

we can see we further need

$$t^* - t^+ > \frac{\mu}{\eta+\mu}\frac{N}{s}$$

The toll price's range can be given by

$$p > \alpha\left[N + \frac{2\mu s}{1+\mu}\left(t^- - t^*\right) - \frac{2\eta s}{1+\mu}\left(t^* - t^+\right)\right]$$
$$\times\left[\frac{\eta(1+\mu)(\mu+\eta) + 2\eta^2\mu}{(1+\mu)(\mu+\eta)}\left(t^* - t^+\right) - \frac{2\eta\mu^2}{(1+\mu)(\mu+\eta)}\left(t^- - t^*\right) - \frac{\eta\mu}{\mu+\eta}\frac{N}{s}\right]$$

and

$$p < \alpha(0)\frac{(1+\mu)\eta}{1+\mu+2\eta}\left(\frac{N}{s} + \frac{2\mu}{1+\mu}t^- - t^+ - \frac{\mu-1}{1+\mu}t^*\right)$$

When $\eta\left(t^* - t^+\right) < \mu\left(t^- - t^*\right)$, we need

$$\tilde{t}_z < t^* + \frac{\eta}{\mu}\left(t^* - t^+\right) \text{ and } \tilde{t}_z < \frac{\eta}{\mu+\eta}\frac{N}{s} + 2t^* - \frac{\eta}{\mu+\eta}t^+ - \frac{\mu}{\mu+\eta}t^-$$

If it holds

$$\frac{\eta}{\mu}\left(t^* - t^+\right) + \frac{\mu}{\eta}\left(t^- - t^*\right) - \frac{N}{s} < 0,$$

we need

$$t^* < \tilde{t}_z < t^* + \frac{\eta}{\mu}\left(t^* - t^+\right)$$

From the plotting of line

$$\frac{\eta}{\mu}\left(t^* - t^+\right) + \frac{\mu}{\eta}\left(t^- - t^*\right) = \frac{N}{s}$$

and line

$$\eta\left(t^* - t^+\right) - \mu\left(t^- - t^*\right) = 0$$

we can see we further need

$$t^* - t^+ < \frac{\mu}{\eta+\mu}\frac{N}{s} \text{ and } t^- - t^* < \frac{\eta}{\mu}\frac{N}{s}$$

The toll price's range can be given by

$$p > \alpha\left[s\frac{\eta+\mu}{\mu}\left(t^{*}-t^{+}\right)\right]\left[\frac{2\eta\mu}{2\eta+1+\mu}\left(t^{-}-t^{*}\right)-\frac{\eta^{2}+\eta\mu+\eta\mu^{2}+3\eta^{2}\mu}{\left(2\eta+1+\mu\right)\mu}\left(t^{*}-t^{+}\right)+\frac{\eta+\eta\mu}{2\eta+1+\mu}\frac{N}{s}\right]$$

and

$$p < \alpha(0)\frac{(1+\mu)\eta}{1+\mu+2\eta}\left(\frac{N}{s}+\frac{2\mu}{1+\mu}t^{-}-t^{+}-\frac{\mu-1}{1+\mu}t^{*}\right)$$

If it holds

$$\frac{\eta}{\mu}\left(t^{*}-t^{+}\right)+\frac{\mu}{\eta}\left(t^{-}-t^{*}\right)-\frac{N}{s} > 0,$$

we need

$$t^{*} < \tilde{t}_{z} < \frac{\eta}{\mu+\eta}\frac{N}{s}+2t^{*}-\frac{\eta}{\mu+\eta}t^{+}-\frac{\mu}{\mu+\eta}t^{-}$$

From the plotting of the lines

$$\frac{\eta}{\mu}\left(t^{*}-t^{+}\right)+\frac{\mu}{\eta}\left(t^{-}-t^{*}\right)=\frac{N}{s}$$

and the line

$$\eta\left(t^{*}-t^{+}\right)-\mu\left(t^{-}-t^{*}\right)=0$$

we can see we further need

$$t^{-}-t^{*} > \frac{\eta}{\mu+\eta}\frac{N}{s}$$

The toll price's range can be given by

$$p > \alpha\left[s\left(t^{*}-t^{+}\right)-s\frac{\mu}{\eta}\left(t^{-}-t^{*}\right)+N\right]\left[\frac{\mu(\eta+2\mu)}{\eta+\mu}\left(t^{-}-t^{*}\right)-\frac{\eta\mu}{\eta+\mu}\left(t^{*}-t^{+}+\frac{N}{s}\right)\right]$$

and

$$p < \alpha(0)\frac{(1+\mu)\eta}{1+\mu+2\eta}\left(\frac{N}{s}+\frac{2\mu}{1+\mu}t^{-}-t^{+}-\frac{\mu-1}{1+\mu}t^{*}\right)$$

When $\eta\left(t^{*}-t^{+}\right)=\mu\left(t^{-}-t^{*}\right)$, the toll price's range can be given by either forms of the unbalanced design. Just notice that the dots $\left(t^{*}-t^{+},t^{-}-t^{*}\right)$ are on the line of $\eta\left(t^{*}-t^{+}\right)-\mu\left(t^{-}-t^{*}\right)=0$.

Appendix C. Lemma 1 proof

Since there are many scenarios related with capacity waste, for simplicity, we only give one scenario's proof as an example. Other scenarios' proof can follow the same logic.

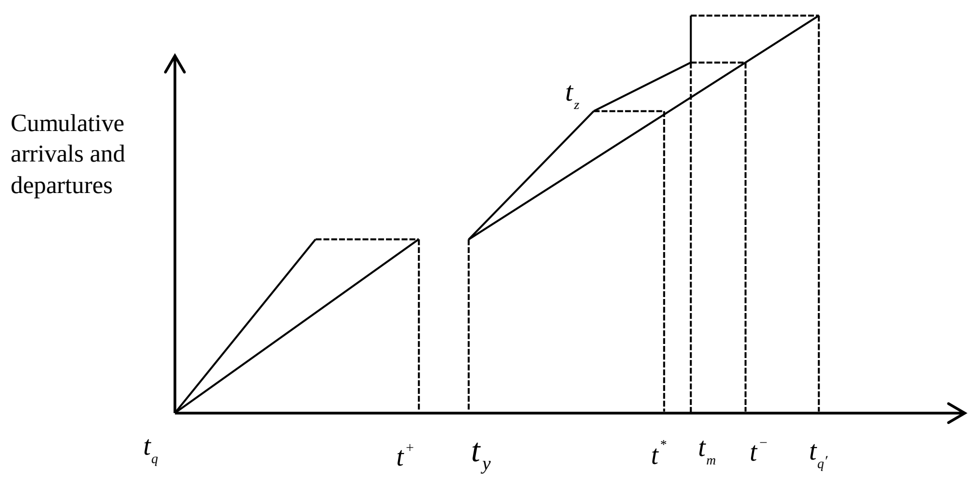

Figure 5: equilibrium profile of toll window only with capacity waste at t^+

For scenario where capacity waste only exists at t^+ (figure 3), from t^+ to t_y, no queue exists at the bottleneck. From the indifferent user's standpoint, the toll price can be obtained as the schedule early delay difference of her coming as the first toll payer and the first toll non-payer, respectively. Thus we have

$$p = \alpha\left(\rho N\right)\eta\left(t_y - t_q\right) \tag{46}$$

From the fact the toll payer has schedule early delay if she is cleared before t^* and has schedule late delay if cleared after t^*, we readily have

$$t^* - t_y = \frac{t_z - t_y}{1-\eta} \tag{47}$$

$$t^- - t^* = \frac{t_m - t_z}{1+\mu} \tag{48}$$

It is easy to obtain

$$\frac{\rho N}{s} = t^- - t_y \tag{49}$$

Based on the definition of equilibrium, the toll non-payer can either arrive before t^+ or join the mass arrival right after the last toll payer arrives. This gives us

$$\eta\left(t^* - t_q\right) = \left(\frac{t^- + t_{q'}}{2} - t_m\right) + \mu\left(\frac{t^- + t_{q'}}{2} - t^*\right) \tag{50}$$

Substituting (49) into (47) gives us

$$t_z = t^* - \eta\left(t^* - t^- + \frac{\rho N}{s}\right) \tag{51}$$

Substituting (51) into (48) gives us

$$t_m = \left(\mu + \eta\right)\left(t^- - t^*\right) + t^- - \frac{\eta \rho N}{s} \tag{52}$$

Substituting (52) into (50) gives us

$$\frac{2\eta + 1 + \mu}{2} t^- - \eta \frac{\rho N}{s} = \frac{\left(1+\mu\right)}{2} t_{q'} + \eta t_q \tag{53}$$

Based on the fact $t_{q'} - t_q = N/s + t_y - t^+$, (49) and (53), we can obtain

$$t^+ = \frac{1+\mu+2\eta}{1+\mu} t_q - \frac{2\eta}{1+\mu} t^- + \frac{2\eta \rho N}{\left(1+\mu\right) s} + \frac{N}{s} - \frac{\rho N}{s} \tag{54}$$

The condition bottleneck only has capacity waste at t^+ is given by

$$t^+ < t_y,$$

which readily gives us

$$t_q < t^- - \frac{\left(2\eta\rho + 1 + \mu\right)N}{\left(1 + \mu + 2\eta\right)s} \tag{55}$$

From (46), (49) and (55), we finally can obtain

$$p > \frac{\alpha\left(\rho N\right)\eta\left(1-\rho\right)\left(1+\mu\right)N}{\left(1+\mu+2\eta\right)s} \tag{56}$$

Based on the fact $t_m \leq t^-$, we can acquire

$$\left(\mu+\eta\right)\left(t^- - t^*\right) + t^- - \frac{\eta\rho N}{s} \leq t^-$$

which readily gives us

$$t^- - t^* \leq \frac{\eta\rho N}{\left(\eta+\mu\right)s} \tag{57}$$

From (49), we can further obtain

$$t^* - t_y \geq \frac{\mu\rho N}{\left(\eta+\mu\right)s} \tag{58}$$

(57) and (58) tell us that

$$\mu\left(t^- - t^*\right) \leq \eta\left(t^* - t_y\right)$$

For a toll window $\left(t^{+\prime}, t^{-\prime}\right)$ designed as $\mu\left(t^{-\prime} - t^*\right) \leq \eta\left(t^* - t^{+\prime}\right)$ with toll price of p_1, we have

$$C\left(x,t\right) = -\beta\left(x\right)\left(t^{-\prime} - t^*\right) + \beta\left(x\right)\frac{\left(1+\mu+2\rho\eta\right)N}{\left(1+\mu+2\eta\right)s}$$

$$C\left(y,t\right) = \alpha\left(y\right)\eta\left(t^* - t^{+\prime}\right) + p_1$$

For a toll window only having capacity waste on t^+, we have

$$C_w\left(x,t\right) = -\beta\left(x\right)\left(t^- - t^*\right) + \beta\left(x\right)\frac{\rho N}{s} + \beta\left(x\right)\frac{p}{\alpha\left(\rho N\right)\eta}$$

$$C_w(y,t) = \beta(y)(t^* - t_y) + p$$

From (56), we can easily obtain

$$C_w(x,t) > -\beta(x)(t^- - t^*) + \beta(x)\frac{(1+\mu+2\rho\eta)N}{(1+\mu+2\eta)s}$$

If we let $t^- = t^{-\prime}$ and choose the same ρ, or namely let $t_y = t^{+\prime}$, we readily have

$$C(x,t) < C_w(x,t) \text{ and } C(y,t) < C_w(y,t)$$

We can see the system cost of the yth user remains same, but under a tolling scheme without capacity waste, she pays less toll. The system cost of the xth user is also lower under a tolling scheme without capacity waste. This tells us for a toll window having capacity waste, we can shorten the toll window to the clearing period of toll payers, then reduce toll price to the critical level corresponding to the amount of the toll payer. With this new toll window, the total system cost is reduced and commuters pay less toll. This proves Lemma 1.

VITA AUCTORIS

Da Xu was born in 1988 in the City of Jinan, China. He graduated from Chung-Ang University of South Korea in 2012, where he obtained a B.B.A. in International Trade and Logistics. He then went to study at University of Windsor. There he obtained an MM in Logistics and Supply Chain Management in 2013 and an M.App.Sc. in Mechanical Engineering in 2015. He is currently a candidate for the Doctoral Degree in Industrial and Manufacturing Systems Engineering at University of Windsor and hopes to graduate in 2018.