Morning commute problem considering route choice, user heterogeneity and alternative system optima

Yang Liu, Yu (Marco) Nie *

Department of Civil and Environmental Engineering, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, United States

A R T I C L E   I N F O

Article history:
Received 28 March 2010
Received in revised form 6 November 2010
Accepted 7 November 2010

Keywords:
Bottleneck model
Congestion pricing
Alternative system optima
User heterogeneity
Social inequity

A B S T R A C T

This paper extends the bottleneck model to study congestion behavior of morning commute and its implications to transportation economics. The proposed model considers simultaneous route and departure time choices of heterogenous users who are distinguished by their valuation of travel time and punctual arrival. Moreover, two dynamic system optima are considered: one minimizes system cost in the unit of monetary value (i.e., the conventional system optimum, or SO) and the other minimizes system cost in the unit of travel time (i.e., the time-based SO, or TSO). Analytical solutions of no-toll equilibrium, SO and TSO are provided and the welfare effects of the corresponding dynamic congestion pricing options are examined, with and without route choice. The analyses suggest that TSO provides a Pareto-improving solution to the social inequity issue associated with SO. Although a TSO toll is generally discriminatory, anonymous TSO tolls do exist under certain circumstances. Unlike in the case with homogenous users, an SO toll generally alters users’ route choices by tolling the poorer users off the more desirable road, which worsens social inequity. Numerical examples are presented to verify analytical results.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

The bottleneck model, proposed by Vickrey (1969) and subsequently refined by Hendrickson and Kocur (1981), Smith (1984), Daganzo (1985), Newell (1987), Arnott et al. (1990), Yang and Huang (1997), among others, has been extensively used in the analysis of congestion behavior of morning commute. In these analyses, commuters must choose their departure times to minimize the sum of travel delays (depending on the bottleneck capacity) and schedule costs (depending on the difference between the actual and desired arrival times). At user equilibrium, no commuter could reduce the total commute cost by unilaterally changing his/her departure time. By endogenizing departure time choice, Vickrey's model explains the pattern of temporal evolution of urban traffic congestion, and has inspired many to develop more realistic extensions for forecasting detailed urban travel patterns under the umbrella of dynamic traffic assignment (DTA). The reader is referred to Peeta and Ziliaskopoulos (2001) for a comprehensive review of the DTA literature.

Although Vickrey's model provides a stunning example in which Pigou's marginal cost toll (Pigou, 1920) completely eliminates congestion at nobody's cost, it does not consider the impacts of user heterogeneity. Pigou’s toll is known to favor individuals with high value of time (VOT), while bringing direct losses to individuals whose value of time savings does not outweigh the paid toll, or those who are “told off” the road and have to use undesirable alternatives (Small, 1983). This regressive incidence of congestion pricing is demonstrated theoretically using both static network analysis (Evans, 1992; Hau, 1998; Yang and Zhang, 2002; Verhoef and Small, 2004) and the bottleneck model (Cohen, 1987; Arnott et al., 1988; Arnott et al., 1994). Cohen (1987) is concerned with two commuter groups: low-income commuters who have lower...
absolute value of time (VOT) and more rigid work schedule (e.g., bus drivers), and high-income commuters with high VOT and flexible work schedule (e.g., stock brokers). Cohen showed that, ignoring the effects of toll revenue redistribution, commuters with low VOT and rigid work schedule are always worse off (at best break even) when Pigou’s toll is implemented. Arnott et al. (1988, 1994) extended Cohen’s analysis by considering other dimensions of user heterogeneity and the impacts of revenue redistribution. In particular, Arnott et al. (1994) showed that, in some cases, the loss of the low-income commuters cannot even be offset by an equal lump-sum refund of toll revenues. Arnott et al. (1994) excludes route choice in their analysis, citing the dominating efficiency gain of rescheduling revealed from an earlier study that considers parallel routes but homogenous users (Arnott et al., 1990). Note that Pigou’s toll does not alter route choice in the homogenous case as shown in Arnott et al. (1990), which probably explains why route choice is less of a concern. Arnott et al. (1992) considered both departure time and route choice using the same heterogeneous model as in Cohen (1987). However, dynamic system optimum was not analyzed because their focus was on time-invariant tolls. Huang (2000) employs a slightly different model to analyze mode choice, which considers two user classes as defined in Cohen (1987), but assumes one of the two parallel links is a transit line and hence has an infinite capacity. Similar to Arnott et al. (1992), however, Huang’s analysis was also limited to time-invariant tolls. Therefore, the first objective of this paper is to characterize dynamic system optima and to reveal the welfare effects of Pigou’s time-varying toll when both departure time and route choices of heterogenous users are considered.

Another issue that has motivated this study is the possibility of considering alternative system optima (SO) in the presence of user heterogeneity. Note that the total system cost can be measured in the unit of either travel time or monetary value, and that minimizing the system cost in either of these units may lead to different SO solutions and pricing schemes. Yang and Huang (2004) and Yang and Huang (2005) examined alternative SO (time-based vs. cost-based) problems using a static multi-class network model. They showed that the cost-based SO flow can be achieved by implementing a marginal cost link toll that is anonymous, i.e., uniform to everyone who uses the same link. However, because the cost-based multi-class SO problem is not generally convex with respect to class-specific link flows, finding the global SO solution can be difficult (Yang and Huang, 2005; Engelson and Lindberg, 2006). In contrast, the time-based SO problem is convex but the marginal cost toll that supports a time-based SO flow is not generally convex with respect to class-specific link flows, finding the global solution can be difficult (Yang and Huang, 2005; Engelson and Lindberg, 2006). In contrast, the time-based SO problem is convex but the marginal cost toll that supports a time-based SO flow is class specific (or discriminatory). Implementing a discriminatory toll scheme is generally difficult because (1) it requires identifying all users’ VOTs, which are not always observable, and (2) it effectively institutes an income redistribution program which easily draws opposition. Yang and Huang (2004) proved that an anonymous (non-marginal-cost) link toll that supports the time-based SO flow always exists and can be obtained by solving a linear program. However, that result only holds when toll can be charged on all links in the network, which may not be satisfied in general, such as when demand is elastic (Clark et al., 2009). Since cost-based and time-based SO tolls demonstrate quite different behavior in the static problem, it is interesting to examine how these differences, such as those dictating welfare effects and practicability of implementation, would manifest in a dynamic setting. Although the overall economical efficiency of a heterogenous transportation system depends on the monetary value of time, minimizing total travel time is a useful alternative objective in its own right, particularly when the focus is to alleviate traffic congestion (note that congestion may be used as a surrogate for mitigating the highly correlated impacts such as emission, fuel consumption, etc.). To the best of our knowledge, few bottleneck-based dynamic morning commute analyses consider time-based SO and the corresponding pricing options.

This paper studies the morning commute problem with heterogenous users using the bottleneck model with two parallel routes. Our setting of user heterogeneity follows Cohen (1987), Arnott et al. (1988) and Arnott et al. (1994), focusing on two groups of users with different VOT and relative cost of schedule delay to travel time. We derive the no-toll equilibrium (NTE) and SO (both conventional SO and time-based SO, or TSO) solutions for the simultaneous choice of routes and departure times. Time-varying tolls to decentralize SO and TSO flow patterns, called SO toll and TSO toll respectively, are examined. Their welfare impacts are analyzed, with and without route choice. Our results indicate that social inequity known to be introduced by SO tolls under user heterogeneity is worsened by route choice, because an SO toll may drive poorer users out of the more desirable routes. On the contrary, TSO tolls are Pareto-improving in that they minimize system travel time without increasing anyone’s commute cost. The existence of anonymous (i.e., the same to anyone who arrives at the same bottleneck at the same time) SO and TSO tolls are also discussed using this conceptual model. We found that the anonymous TSO toll may not exist even when the demand is fixed.

The rest of the paper is organized as follows. Section 2 introduces the basic model setting and makes necessary definitions. Section 3 revisits the single-route problem and introduces the new time-based SO analysis. The results in Section 3 are extended to the two-route problem in Section 4, which reveal how SO and TSO lead to different route choice patterns and welfare effects. Numerical examples are provided in Section 5 to verify the analytical results, and Section 6 summarizes our findings and sets out directions for future research.

2. Model setting

Consider a network with one origin–destination (O–D) pair connected by two parallel routes: a highway (denoted as route 1), and a local street (denoted as route 2). Let $T_j$ and $s_j$ be the free flow travel time and bottleneck capacity on route

---

1 Cost-based in Yang and Huang (2005) refers to measuring system cost in the unit of monetary value. In this paper, their cost-based SO is termed conventional SO, and referred to using the acronym “SO”.
can be written mathematically as

\[ \text{total system cost, let } x \]

To simplify the analysis, it is also assumed that both groups have the same relative cost of late to early arrival i.e.,

\[ \eta = \frac{\gamma_i}{\beta_i}, \quad i = b, w \]

Moreover, the following identities will be used in our analysis.

\[ \phi = \frac{\eta}{1 + \eta}, \quad \theta_i = \frac{x_i}{x_i - \beta_i}, \quad \rho_i = \frac{x_i}{\gamma_i + x_i}, \quad \delta_i = \frac{\beta_i \gamma_i}{\beta_i + \gamma_i} = \phi \frac{\beta_i}{x_i} = 1 - \frac{1}{\theta_i}, \quad i = w, b \]

A list of notation can be found in Appendix A.

3. Single-route problem

For the sake of completeness, this section first reviews the no-toll equilibrium (NTE) and system optimum (SO) solutions for the single-route bottleneck problem with two classes of users (w, b). For simplicity, we shall denote the capacity of the single bottleneck as s and set the free flow travel time as zero.

3.1. No-toll equilibrium

At NTE, the commuters in each group must have the same commute cost. In this case, group b should depart at the center of the rush hour because its relative cost of schedule delay to travel time is higher, whereas group w departs on the tails, i.e., before or after group b (Cohen, 1987; Arnott et al., 1994). Since there is no overlap between the departures of the two groups and noting that \( \theta_w < \theta_b, \rho_w > \rho_b \) by Assumption (1), the NTE solution for the two-class single-route problem can be depicted in Fig. 1. The commute costs of group w and b at NTE are given by

\[ c_w = \frac{\delta_w (N_w + N_b)}{s} \]

\[ c_b = \frac{x_b \delta_w N_w}{s} + \frac{\delta_b N_b}{s} \]

3.2. System optimum

At the system optimum, no queueing delays should be present and therefore the departure and arrival curves overlap, see Fig. 2(a). Since schedule delay is the only cost involved in travel, the departing order depends on the value of schedule penalty parameters \( \beta_i \) and \( \gamma_i \).

3.2.1. \( \beta_w > \beta_b \)

When \( \beta_w > \beta_b \), group w should travel at the center of the rush hour in order to minimize the total schedule delay. To obtain total system cost, let \( x_1 = \gamma - t_1, y_1 = t_2 - t \), \( x_2 = t_1 - t_0, y_2 = t_3 - t_2 \) (cf. Fig. 2, where \( t_1 \) and \( t_2 \) are the first and last departure epochs of white-collar group). The total schedule cost of group w, denoted as \( C_w \), is

\[ C_w = 0.5x_1^2 \beta_w s + 0.5y_1^2 \gamma_w s \]

where \( x_1 + y_1 = N_w s \). The total schedule cost of group b, denoted as \( C_b \), is

\[ C_b = 0.5(x_2^2 + 2x_1 x_2) \beta_b s + 0.5(y_2^2 + 2y_1 y_2) \gamma_b s \]

\[ 2 \quad \text{This relationship was empirically verified in a study conducted for the San Francisco Bay Area (Small, 1982).} \]

\[ 3 \quad \text{The results are given in Arnott et al. (1988). For the convenience of the reader, a derivation of the results is provided in Appendix B.} \]
where $x_2 + y_2 = N_b/s$. Let $C = C_w + C_b$, minimizing $C$ with respect to $x_i$, $i = 1, 2$ requires

$$\frac{\partial C}{\partial x_1} = x_1(b_w + \gamma_w)s + x_2(b_b + \gamma_b)s - N_b\gamma_b - N_w\gamma_w = 0$$

$$\frac{\partial C}{\partial x_2} = x_1(b_b + \gamma_b)s + x_2(b_b + \gamma_b)s - N_b\gamma_b - N_w\gamma_b = 0$$

Subtracting the first equation from the second yields

$$x_1(\beta_b + \gamma_b - \beta_w - \gamma_w) = N_w(\gamma_b - \gamma_w)$$

Since $\gamma = \eta \beta$, the above can be rewritten as

$$x_1(1 + \eta)(\beta_b - \beta_w) = N_w\eta(\beta_b - \beta_w)$$

Thus, we obtain

$$\begin{align*}
x_1 &= \frac{\eta N_w}{1 + \eta}, & y_1 &= \frac{(1 - \phi)N_w}{s}, \\
x_2 &= \frac{\phi N_b}{s}, & y_2 &= \frac{(1 - \phi)N_b}{s}
\end{align*}$$

(6)

Plugging $x_1$, $x_2$, $y_1$, $y_2$ into $C_w$ and $C_b$ (noting that $\phi^2 + (1 - \phi)^2 = 1$), we obtain the minimum system cost as

$$C^* = \frac{\phi}{2s} \left( \beta_w N_w^2 + \beta_b N_b^2 + 2\beta_b N_w N_b \right)$$

(7)

Thus, the marginal cost of group $w$ and $b$, i.e., the increase of system cost corresponding to an additional user from group $w$ (or group $b$), is...
the center of the rush hour. Repeating the above analysis we have

\[ m_{cw} = \frac{\partial C}{\partial Nw} = \frac{\delta w Nw}{s} + \frac{\delta b Nb}{s} \]  
(8)

\[ m_{cb} = \frac{\partial C}{\partial Nb} = \frac{\delta b (Nw + Nb)}{s} \]  
(9)

Let \( g(t) \) be the schedule cost of the commuter departing at time \( t \), \( g(t) \) can be written as

\[
 g(t) = \begin{cases} 
 \beta_b (t' - t), & \text{for } t \in [t_0, t_1] \\
 \beta_w (t' - t), & \text{for } t \in (t_1, t') \\
 \gamma_w (t - t'), & \text{for } t \in (t', t_2) \\
 \gamma_b (t - t'), & \text{for } t \in (t_2, t_1] 
\end{cases} 
\]  
(10)

where one can solve from Eq. (6) that

\[
 t_0 = t' - \frac{\phi N_b + N_w}{s}, \quad t_1 = t' - \frac{\phi N_w}{s}, \quad t_2 = t' + (1 - \phi) \frac{N_w}{s}, \quad t_3 = t' + (1 - \phi) \frac{N_b + N_w}{s} 
\]  
(11)

Note that the marginal cost toll at departure time \( t \), denoted as \( \tau(t) \), can be defined as the difference between the marginal cost \( mc \) and the schedule cost \( g(t) \), that is,

\[
 \tau(t) = \begin{cases} 
 mc - g(t), & t \in [t_0, t_1] \\
 mc_b - g(t), & t \in (t_1, t_2] 
\end{cases} 
\]  
(12)

Substituting \( mc_b \) and \( g(t) \) in the above equation with (8)–(10), for \( t \in [t_0, t_1] \), we obtain

\[
 mc_b - g(t) = \frac{\delta b (Nw + Nb)}{s} - \beta_b (t' - t_0 - (t - t_0)) = \frac{\delta b (Nw + Nb)}{s} - \beta_b \phi \frac{N_b + N_w}{s} + \beta_b (t - t_0) 
\]

Since \( \beta_b \phi = \delta_b \) by (3), \( mc_b - g(t) = \beta_b (t - t_0) \). Proceeding with other intervals in a similar fashion, we have

\[
 \tau(t) = \begin{cases} 
 \beta_b (t - t_0), & t \in [t_0, t_1] \\
 \beta_w (t_1 - t_0) + \beta_b (t - t_1), & t \in (t_1, t') \\
 \gamma_w (t_2 - t_1) + \gamma_b (t_2 - t), & t \in (t', t_2] \\
 \gamma_b (t_2 - t), & t \in (t_2, t_1] 
\end{cases} 
\]  
(13)

The above toll scheme is depicted in Fig. 2(b).

3.2.2. \( \beta_w < \beta_b \)

When \( \beta_w < \beta_b \) the departure order will be the same as in NTE case, i.e., group \( w \) travels on the tails while group \( b \) travels at the center of the rush hour. Repeating the above analysis we have

\[
 m_{cw} = \frac{\delta w (Nw + Nb)}{s} 
\]  
(14)

\[
 m_{cb} = \frac{\delta w Nw + \delta b Nb}{s} 
\]  
(15)

\[
 g(t) = \begin{cases} 
 \beta_w (t' - t), & t \in [t_0, t_1] \\
 \beta_b (t' - t), & t \in (t_1, t') \\
 \gamma_w (t - t'), & t \in (t', t_2) \\
 \gamma_b (t - t'), & t \in (t_2, t_1] 
\end{cases} 
\]  
(16)

\[
 t_0 = t' - \frac{\phi N_b + N_w}{s}, \quad t_1 = t' - \frac{\phi N_w}{s}, \quad t_2 = t' + (1 - \phi) \frac{N_w}{s}, \quad t_3 = t' + (1 - \phi) \frac{N_b + N_w}{s} 
\]  
(17)

\[
 \tau(t) = \begin{cases} 
 mc_w - g(t), & t \in [t_0, t_1] \\
 mc_b - g(t), & t \in (t_1, t_2] 
\end{cases} 
\]  
(18)

3.2.3. Welfare effects of the marginal SO toll

Unlike in the bottleneck model with homogenous users, the marginal cost at SO in both cases is not necessarily equal to the cost at NTE. Specifically, group \( w \) is either better off (\( \beta_w > \beta_b \) and the departure order is reversed) or breaks even (\( \beta_w < \beta_b \) and the departure order is preserved), while group \( b \) is always worse off. To see this, note that \( mc_w - c_w = 0 \) when \( \beta_w < \beta_b \) but
We emphasize that the SO toll discussed above is anonymous, in that it can be implemented without differentiating users. In other words, the toll is same as long as an individual arrives at the bottleneck at the same time, regardless of to which group it belongs.

3.3. Time-based system optimum

The SO model discussed above minimizes the total monetary value of system cost, following most studies in the literature (e.g. Cohen, 1987; Arnott et al., 1994). It has been shown (Yang and Huang, 2004), for the static multiple-class network model, that the system optimum solutions and the corresponding pricing schemes differ when the system cost is measured in monetary and time units. In what follows, Yang and Huang’s result is extended to the single-route bottleneck case.

At the time-based SO (TSO), each group’s schedule delay is valued by $\beta_i / z_{ti}$, which dictates the departure order. According to Assumption (1), since group $b$ has a larger $\beta_i / z_{ti}$ ratio, it travels at the center of the rush hour while group $w$ travels at the tails, as they do at NTE. Because all queuing delays have to be removed at TSO, the total system cost would consist of only schedule costs, and in the unit of travel time it can be written as

$$TT = \{0.5x_1^2 + 0.5y_2^2 + 0.5(x_1^2 + 2x_1y_2)\beta_{w} s + 0.5(y_2^2 + 2y_1y_2)\gamma_{w} s\}/z_{tw},$$

where $x_i, y_i, i = 1, 2$ are defined in Eq. (6). Following the same analysis, the reader can verify the optimal values of $x_1, x_2, y_1, y_2$ when $TT$ is minimized as

$$x_1 = \frac{\eta}{1 + \eta} \frac{N_b}{s} = \frac{\phi N_b}{s}, \quad y_1 = \frac{(1 - \phi)N_b}{s}, \quad x_2 = \frac{\phi N_w}{s}, \quad y_2 = \frac{(1 - \phi)N_w}{s} \quad \text{(19)}$$

Accordingly, the optimal system travel time can be obtained as

$$TT^* = \frac{\phi}{2s} \left(\beta_{w} N_w^2 / z_{tw} + \beta_{b} N_b^2 / z_{tw} + 2\beta_{w} N_w N_b / z_{tw}\right) \quad \text{(20)}$$

The marginal travel delay for both groups are

$$mt_w = \frac{\partial TT}{\partial N_w} = \frac{\delta_s (N_w + N_b)}{s z_{tw}} \quad \text{(21)}$$

$$mt_b = \frac{\partial TT}{\partial N_b} = \frac{\delta_s N_b + \delta_s N_w}{s z_{tw}} \quad \text{(22)}$$

Accordingly, the commute cost for each group at TSO is given as

$$mc_w = mt_w z_{tw} = \frac{\delta_s (N_w + N_b)}{s} \quad \text{(23)}$$

$$mc_b = mt_b z_{tw} = \frac{\delta_s N_b + \delta_s N_w}{s z_{tw}} \quad \text{(24)}$$

Interestingly, at TSO, the marginal cost equals the commute cost at NTE (Eqs. (4) and (5)). That is, since the tolls that decen-tralize TSO replace the cost of queuing delay at NTE, everyone breaks even at the tolled equilibrium just as in the homogeneous case. However, unlike the conventional SO, the TSO toll may be discriminatory, as shown below.

Using Eqs. (23), (24), (16), (17) and the first half of (18), the marginal TSO toll can be obtained as follows:

$$\tau(t) = \left\{\begin{array}{ll}
\beta_w (t - t_0), & t \in [t_0, t_1] \\
\frac{\delta_s N_w}{s z_{tw}} + \beta_b (t - t_1), & t \in (t_1, t^*) \\
\frac{\delta_s N_w}{s z_{tw}} + \gamma_w (t - t_1), & t \in [t_1, t^*] \\
\gamma_w (t - t_1), & t \in (t_2, t_3]
\end{array}\right. \quad \text{(25)}$$

Note that group $w$ travels when $t \in [t_0, t_1] \cup (t_2, t_3]$ and group $b$ travels when $t \in (t_1, t_2]$. It is worth emphasizing that the above toll function is not continuous at $t_1$ and $t_2$. To see this, note that at $t_1$ and $t_2$, the tolls for group $w$ and $b$ are $\frac{\delta_s N_w}{s z_{tw}}$ (i.e., $\beta_w (t_1 - t_0)$, where $t_0, t_1$ are defined in Eq. (17)) and $\frac{\delta_s N_b}{s z_{tw}}$, respectively. Fig. 3 visualizes the above marginal TSO toll when $\beta_w > \beta_b$ and otherwise.

We first discuss the case when $\beta_w > \beta_b$, see Fig. 3(a). Note that group $w$ pays a higher toll at $t_1$ and $t_2$ since

$$\frac{\delta_s N_w}{s z_{tw}} < \frac{\delta_s N_b}{s}$$

by Assumption (1). Consequently, discontinuities arise at $t_1$ and $t_2$: from $t_0$ to $t_1$ and then from $t_2$ to $t_3$, the toll for group $b$ is depicted by the bold black line; whereas between $t_1$ to $t_2$, the blue dashed line depicts the toll for group $b$. It is important to notice that the toll for group $b$ (described by the blue dash line) is not enough to offset the group $w$’s gains from the reduction of schedule cost as the arrival time moves closer to $t^*$. To see this, note that the red dot line, along which group $w$’s schedule
delay cost is compensated by toll, always lies above the blue dash line between $t_1$ and $t_2$. In other words, if group $w$ pays the same toll as group $b$ does between $t_1$ and $t_2$, these white-collar workers can lower the commute cost by managing to arrive during that period. Therefore, the toll has to be set at different levels for groups $b$ and $w$ in order to support the TSO flow pattern. In contrast, on the tails, an anonymous toll equal to the toll of group $w$ is large enough to keep group $b$ from leaving the middle slot. Thus, the toll on the tails is anonymous.

Similar analysis can be performed for the case of $b_w < b_b$, in which TSO produces the exactly same flow pattern as SO. As shown in Fig. 3(b), an anonymous toll is sufficient for the period of $[t_0, t_1] \cup [t_2, t_3]$ (the solid black line). In the middle of the rush hour, the discriminatory toll is needed to maintain a stable SO flow pattern wherever the toll required to balance the schedule cost of group $w$ is larger than the designated toll of group $b$ (i.e., when the blue dash line lies below the red dot line).

Tolls that can decentralize a TSO flow pattern are not unique. Fig. 4(a) demonstrates a class of alternative TSO toll schemes which take the form

$$
\tau(t) = \begin{cases} 
\beta_w(t - t_0), & t \in [t_0, t_1] \\
\tau^0_b + \beta_b(t - t_1), & t \in (t_1, t^*] \\
\tau^0_b + \gamma_b(t_2 - t), & t \in (t^*, t_2] \\
\gamma_w(t_3 - t), & t \in (t_2, t_3] 
\end{cases}
$$

Note that all toll schemes satisfying (26) require discriminatory toll between $t_1$ and $t_2$ to attain TSO. However, anonymous toll on the tails is always ensured by the upper bound set on $\tau^0_b$. For example, when the toll on group $b$ at $t_1$ is less than or equal to $\delta_b N_w/s$, the same anonymous toll, depicted by the black curve in Fig. 4(a), is always larger than the toll required to
balance the increase in schedule cost of group $b$; otherwise, a discriminatory toll scheme would be needed on the tails. Interestingly, when $\beta_w < \beta_b$, an anonymous toll scheme for the entire rush hour can be obtained $t^{0,b}_w = \delta_s N_w / s$. As shown in Fig. 4(b), the red curve which presents the toll for group $w$ is no longer needed in this case as it never lie above the toll curve for group $b$ for any $t \in [t_1, t_2]$.

We summarize the results from this section in Table 1. First, our analysis shows that all travelers break even at the marginal TSO toll, while the conventional SO toll is generally in favor of those with higher value of time yet a flexible work schedule at the expense of the other group, which raises the equity issue. The analysis also indicates that the marginal TSO toll cannot be anonymous in general, which agrees with the findings of Yang and Huang (2004) for the static problem. Moreover, an anonymous toll that decentralizes TSO may not always exist, as illustrated in the case of $\beta_w > \beta_b$. This is consistent with the finding for the static alternative-criteria congestion pricing problem (Clark et al., 2009).

![Alternative toll schemes that decentralize the time-based SO solution.](image)

---

**Table 1**

A summary of the single-route bottleneck problem with heterogeneous users (with Assumption (1)).

<table>
<thead>
<tr>
<th>Departure order</th>
<th>Anonymous toll exists?</th>
<th>Welfare effects of marginal toll</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_w &gt; \beta_b$</td>
<td>$\beta_w &lt; \beta_b$</td>
</tr>
<tr>
<td>NTE</td>
<td>w, b, w</td>
<td>N/A</td>
</tr>
<tr>
<td>SO</td>
<td>b, w, b</td>
<td>w, b, w</td>
</tr>
<tr>
<td>TSO</td>
<td>w, b, w</td>
<td>w, b, w</td>
</tr>
</tbody>
</table>
4. Two-route problem

4.1. No-toll equilibrium

For the network with two parallel routes, the free flow travel times on both routes have to be taken into consideration because they may affect route choice. Recalling Eqs. (4) and (5), the travel cost on route \( j \) at NTE can be written as:

\[
\begin{align*}
\text{where } d_i & \text{ is defined in Eq. (3); } c_j^i \text{ represents the individual cost of class } i \text{ on route } j; T_j \text{ is the free flow travel time on route } j \text{ (} T_1 < T_2 \text{); and } f_j^i \text{ represents the flow of class } i \text{ on route } j. \\
\end{align*}
\]

The equilibrium condition implies that within the same group the cost on used route is lower than or equal to the cost of unused one, i.e.,

\[
f_j^i = 0 \Rightarrow c_j^i \geq c_j^i(k \neq j), \forall i.
\]

To simplify the notation, let us define

\[
N'/C^3 = s_1 a b d b \frac{T_2 - T_1}{C^0}, \quad N'' = s_1 \frac{N_w}{C^3} (T_2 - T_1),
\]

where \( N'' > N' \) by Assumption (1). We have the following result regarding the flow distribution between the two routes.

**Theorem 1.** For the two-route two-class bottleneck model with Assumption (1), one of the following flow distribution schemes may arise at NTE depending on the demand pattern.

1. When

\[
N_w + \frac{a_w}{a_b} N_b \leq N''
\]

all users will stay on route 1.

2. When

\[
N_w + \frac{a_w}{a_b} N_b > N''
\]

and \( N_w \leq N'' \)

group \( w \) only uses route 1 and group \( b \) is distributed on both routes according to

\[
\begin{align*}
f^1_w & = \frac{1}{s_1 + s_2} \left[ -s_2 \frac{\frac{a_w}{a_b} N_w + s_1 N_b + s_2 N''}{\frac{a_w}{a_b}} \right] \\
f^2_w & = \frac{1}{s_1 + s_2} \left[ s_2 \frac{\frac{a_w}{a_b} N_w + s_2 N_b - s_2 N''}{\frac{a_w}{a_b}} \right]
\end{align*}
\]

3. When \( N_w \geq N'' \), group \( w \) and \( b \) are distributed on both routes according to

\[
\begin{align*}
f^1_w & = N'' + \frac{s_1}{s_1 + s_2} (N_w - N'') \\
f^2_w & = \frac{s_2}{s_1 + s_2} (N_w - N'') \\
f^1_b & = N_b \frac{s_1}{s_1 + s_2} \\
f^2_b & = N_b \frac{s_2}{s_1 + s_2}
\end{align*}
\]

**Proof**

(1) When demand is low, all users will choose route 1 because of the assumption \( T_1 < T_2 \). In fact, they will continue to stay on route 1 when

---

\footnote{\textsuperscript{4} Arnott et al. (1992) obtained similar results for this case. Particularly, Eqs. (35)-(38) are identical to Eq. (27) and (28) in Arnott et al. (1992).}
We claim that \( c^1_b \leq x_b T_2 \), \( i = w, b \).

We claim that \( c^1_i \leq x_i T_2 \rightarrow c^1_i \leq x_w T_2 \). To see this, note that

\[
\frac{c^1_i}{x_i} - \frac{c^1_b}{x_b} = \left( \frac{\delta_i}{x_i} - \frac{\delta_b}{x_b} \right) \frac{N_b}{s_1} = \left( \frac{\beta_i}{x_i} - \frac{\beta_b}{x_b} \right) \phi N_b < 0
\]

where the first equality follows from Eqs. (27) and (28), the second equality follows from the identity (3), and the inequality follows from the assumption (1). Thus, all users will select route 1 so long as

\[
c^1_i \leq x_i T_2 \iff N_w + \frac{\delta_i}{x_i} \frac{N_b}{s_1} \leq s_1 \frac{\delta_i}{x_i} (T_2 - T_1) = N^w
\]

which proves Case 1.

(2) The proof of (1) implies that group \( b \) will start to use route 2 when the demand violates condition (30). To keep group \( w \) from route 2 requires that it is cheaper for group \( w \) to stay on route 1. Let’s first solve the flow distribution of group \( b \) \((f^1_b, f^2_b)\) according to the UE conditions:

\[
\begin{align*}
    c^1_b &= c^2_b \\
    f^1_b + f^2_b &= N_b
\end{align*}
\]

Plugging the NTE cost (28) into the above equation system with \( f^1_w = N_w, f^2_w = 0 \), we can solve \( f^b \) as given in Eqs. (33) and (34). The cost of group \( w \) on route 1 is:

\[
c^1_w = \delta_w \frac{N_w + f^1_w}{s_1} + x_w T_1
\]

Now, suppose a group \( w \) user would switch to route 2, she/he would face a cost

\[
c^w_w = \delta_w \frac{f^2_w}{s_2} + x_w T_2
\]

However, the above switch should not happen if \( c^1_w \leq c^w_w \). The reader can verify that this inequality implies that when \( N_w \leq N^\ast \), group \( w \) will not use route 2.

(3) As shown above, when \( N_w > N^\ast \), group \( w \) starts to use route 2. In this case, the flow distribution of both groups as shown in Eqs. (35)–(38) can be solved from the UE conditions:

\[
\begin{align*}
    c^1_i &= c^2_i, \quad c^1_w &= c^2_w \\
    f^1_b + f^2_b &= N_b, \quad f^1_w + f^2_w &= N_w
\end{align*}
\]

This completes the proof. □

Fig. 5 visualizes the route choice pattern as a function of \( N_w \) and \( N_b \). As illustrated in the figure, group \( w \) tends to have a higher ratio of the faster route (route 1) users than group \( b \). This is expected because group \( w \) has a higher value of time. What is less expected is the fact that the route choice of group \( w \) is completely independent of the presence of group \( b \). Specifically, when \( N_w < N^\ast \), group \( w \) only uses route 1; otherwise, the extra flow \( N_w - N^\ast \) will be distributed among two
routes proportional to bottleneck capacity. These decisions have nothing to do with \( N_b \). This independence suggests the superior position of group \( w \) in the game. On the other hand, the route choice of group \( b \) is largely affected by the presence of group \( w \). When \( N_b < N' \), for instance, group \( b \) could end up using route 1 only, or distributing over both routes with various ratios, depending on how many group \( w \) users may use the network during the rush hour.

Another observation from the figure is that group \( b \) always gets a fair share on the fast route. Even in the worse case \( (N_w > N'), s_1/(s_1 + s_2) \) group \( b \) users will still be able to stay on the fast route.

### 4.2. System optimum

We now extend our analysis in Section 3.2 to the two-route problem. To avoid undue complexities, hereafter we further assume \( \beta_w > \beta_b \). This case concerns us more because it reverses the departure slot at NTE, which turns out to worsen the inequality issue. With this additional assumption in mind and recalling Eq. (7), the SO route choice pattern can be obtained by solving the following minimization problem.

\[
\min C(f) = \sum_i \sum_j a_i^f T_j + \sum_j \frac{1}{2s_j} \left[ 2\delta_b f_b^j f_b^2 + \delta_b (f_b^j)^2 + \delta_w (f_w^j)^2 \right]
\]

s.t. \[
\sum_j f_j^i = N_i, \quad \text{for } i = w, b \tag{39}
\]

\[
f_j^i \geq 0, \quad \text{for } i = w, b; j = 1, 2 \tag{40}
\]

The KKT conditions of the above problem include the feasible conditions (39) and (40) as well as the following complementarity conditions

\[
f_b^j (mc_b^j - \mu_b) = 0, \quad mc_b^j - \mu_b \geq 0 \quad \text{for } j = 1, 2 \tag{41}
\]

\[
f_w^j (mc_w^j - \mu_w) = 0, \quad mc_w^j - \mu_w \geq 0 \quad \text{for } j = 1, 2 \tag{42}
\]

where \( \mu_i, i = w, b \) are multipliers associated with constraints (39) and

\[
mc_w^j = a_w T_j + \frac{1}{s_j} (\delta_b f_b^j + \delta_w f_w^j), \quad mc_b^j = a_b T_j + \frac{1}{s_j} (\delta_b f_b^j + \delta_b f_b^j) \tag{43}
\]

denote the marginal cost of groups \( w \) and \( b \), respectively, with respect to usage on route \( j \). The objective function \( C(f) \) is strictly convex. To see this, note that its Hessian

\[
H_c = \begin{bmatrix}
H_{11} & 0 \\
0 & H_{22}
\end{bmatrix}, \quad \text{where } H_{11} = \frac{1}{s_1} \begin{bmatrix}
\delta_w & \delta_b \\
\delta_b & \delta_b
\end{bmatrix}, \quad H_{22} = \frac{1}{s_2} \begin{bmatrix}
\delta_w & \delta_b \\
\delta_b & \delta_b
\end{bmatrix}
\]

\( H_c \) is evidently positive definite since \( \delta_w > \delta_b \iff \phi \beta_w > \phi \beta_b \) (and \( \beta_w > \beta_b \) is assumed to hold in this section). Consequently, KKT conditions are also sufficient to ensure a unique global optimal solution.

To simplify the notation, we further introduce \( N'_w \) and \( N'_b \) as:

\[
N'_w = s_1 \frac{2\beta_w - \beta_b}{\delta_w - \delta_b} (T_2 - T_1), \quad N'_b = s_2 \frac{2\beta_b}{\beta_b} (T_2 - T_1) \tag{44}
\]

\[
N'_w = s_1 (\frac{2\beta_w}{\delta_w} - \frac{2\beta_b}{\delta_b}) (T_2 - T_1), \quad N'_b = s_2 \frac{2\beta_b}{\delta_b} (T_2 - T_1) \tag{45}
\]

**Theorem 2.** For the two-route two-class bottleneck model, one of the following five flow distribution schemes may arise at SO depending on the demand pattern.

a. When
\[
N_w + N_b \leq N',
\]
all the users stay on route 1, so route 2 remains unused.

b. When
\[
N_w + N_b > N',
\]
\[
N_w \leq N'_w \tag{46}
\]
and \( N_w - \frac{s_1}{s_2} N_b < N' \) (and \( N_w - \frac{s_1}{s_2} N_b < N' \))

\[
N_w > N'_w \tag{47}
\]

\[
N_w \geq N'_w \tag{48}
\]

\[
N_w - \frac{s_1}{s_2} N_b < N' \tag{49}
\]

\( \text{group } w \) only uses route 1, and group \( b \) is assigned to both routes according to
(b) Case b implies that

\[ f_b^1 = \frac{1}{s_1 + s_2} \left[ -s_2 N_w + s_1 N_b + s_2 N' \right] \]  
\[ f_b^2 = \frac{1}{s_1 + s_2} \left[ s_2 N_w - s_2 N' \right] \]  

c. When

\[ N_w > N'_w \text{ and } N_b > N'_b \]  

both groups' users are assigned to the two routes. The optimal flow distributions are:

\[ f_w^1 = \frac{1}{s_1 + s_2} \left[ s_1 N_w + s_1 N'_w \right] \]  
\[ f_w^2 = \frac{1}{s_1 + s_2} \left[ s_1 N_w - s_2 N'_w \right] \]  
\[ f_b^1 = \frac{1}{s_1 + s_2} \left[ s_1 N_b - s_1 N'_b \right] \]  
\[ f_b^2 = \frac{1}{s_1 + s_2} \left[ s_2 N_b + s_1 N'_b \right] \]  

d. When

\[ N_w - \frac{s_1}{s_2} N_b \geq N' \]  
\[ N_w - \frac{s_1}{s_2} N_b \leq N'' \]  

group w only uses route 1 and group b only uses route 2.

e. When

\[ N_w - \frac{s_1}{s_2} N_b > N'' \]  
\[ N_b \leq N'_b \]  

group b only uses route 2 and group w is distributed to both routes according to:

\[ f_w^1 = \frac{1}{s_1 + s_2} \left[ s_1 N_w + s_1 \frac{\partial b}{\partial w} N_b + s_2 N'' \right] \]  
\[ f_w^2 = \frac{1}{s_1 + s_2} \left[ s_2 N_w - s_1 \frac{\partial b}{\partial w} N_b - s_2 N'' \right] \]  

Proof

(a) In this case, when all demands are assigned to route 1, the marginal costs on route 1 must still be smaller than those on route 2 for both groups. For group b, this implies

\[ mc_b^1 = \alpha_b T_1 + \frac{1}{s_1} (\delta_N N_w + \delta_b N_b) \leq mc_b^2 = \alpha_b T_2 \iff N_w + N_b \leq s_1 \frac{\partial b}{\partial b} (T_2 - T_1) = N' \]

which gives rise to Condition (46). Also note that Condition (46) is sufficient to guarantee that \( mc_b^1 < mc_b^2 \) here. To see this, note that

\[ N_w + N_b \leq s_1 \frac{2b}{\partial b} (T_2 - T_1) \leq s_1 \frac{\partial w}{\partial w} (T_2 - T_1) \]

because \( \alpha_b / \delta_b < \alpha_w / \delta_w \), which leads to

\[ \delta w N_w + \delta w N_b \leq s_1 \alpha_w (T_2 - T_1) \iff \alpha_w T_1 + \frac{1}{s_1} (\delta w N_w + \delta w N_b) = mc_w^1 \leq \alpha_w T_2 = mc_w^2 \]

(b) Case b implies that \( f_w^1 = N_w, f_b^2 = 0 \) and \( f_b^1 > 0 \). Thus, the flow distribution for the blue-collar group can be solved from

\[ mc_b^1 = mc_b^2, \quad f_b^1 + f_b^2 = N_b \]

as given in Eqs. (50) and (51). Conditions (49) and (47) follow from \( f_b^1 > 0 \) and \( f_b^2 > 0 \), respectively. Moreover, since group w does not use route 2, the KKT condition requires \( mc_w^1 = mc_w^2 \), which leads to Condition (48).
(c) Both routes are used by the two groups in this case. Consequently, the flow distribution of both groups can be solved from

\[ mc_i^l = mc_i^2, \quad f_i^1 + f_i^2 = N_i, \quad i = b, w \]

as shown in Eqs. (53)–(56). Then, \( N_w > N_b^* \) follows from \( f_w^2 > 0 \) and \( N_b > N_b^* \) follows from \( f_b^1 > 0 \).

(d) Because group \( b \) only uses route 2, the KKT conditions suggest that \( mc_i^2 \leq mc_i^b \), which is reduced to Condition (57). Similarly, the fact that group \( w \) only uses the route 1 indicates \( mc_i^1 \leq mc_i^w \), which leads to Condition (58).

(e) In this case, we solve the flow distribution of group \( w \) from (with \( f_b^1 = 0, f_w^2 = N_b^* \))

\[ mc_i^w, f_w^2 + f_w^2 = N_w \]

as given in Eqs. (61) and (62). Subsequently, \( f_w^2 > 0 \) gives condition (59). Moreover, since group \( b \) only uses route 2, it follows that \( mc_i^2 \geq mc_i^b \), which can be simplified as Condition (60).

This completes the proof. \( \square \)

The above five cases at SO are illustrated in Fig. 6. Note that cases d and e represent two new distribution patterns that do not exist at NTE. In these two cases, the demand of group \( b \) is low and the demand of group \( w \) is high, and group \( b \) is assigned to the slow route (route 2) while the group \( w \) either stays on the fast route (route 1) or splits between the two routes. Interestingly, Case d represents a complete segmentation, i.e., group \( w \) only uses the fast route and group \( b \) only uses the slow route.

Fig. 7 combines boundaries from Figs. 5 and 6 to delineate nine NTE-SO transition subareas. Analyzing these subareas allows one to track how the flow distributions vary from NTE to SO. For example, Subarea 1-a admits Case 1 at NTE and Case a at SO. In Fig. 7, the two blue dash lines represent the boundary conditions (30) and (32) at NTE, and the five red solid lines represent the boundary conditions (46), (48), (57), (58) and (60) at SO.

Theorem 2 suggests that the route choices are different at NTE and SO. In other words, the route usage at NTE is no longer system optimal in the presence of user heterogeneity. The following result reveals how class-specific flows and total flows on route 2 are changed from NTE to SO.

**Theorem 3.** If Assumption (1) is satisfied and \( \beta_w > \beta_b \), then

\[ f_b^2 > f_b^2, \quad f_w^2 < f_w^2, \quad f_w^2 + f_w^2 > f_b^2 + f_b^2 \]

![Fig. 6. Five cases at SO for the two-route problem.](image-url)
where $\hat{f}_j^i$ and $\tilde{f}_j^i$ represent flow of group $i$ on route $j$ at SO and NTE, respectively. Moreover, the commute cost of group $w$ (b) is always lower (higher) at SO than at NTE.

**Proof.** The second and third columns in Table 2 list the changes of $f_2^b$ and $f_2^w$ in all nine subareas, which are obtained by inspecting the NTE and SO solutions corresponding to each subarea. From these results, we can clearly see that $f_2^b \geq \hat{f}_2^b$ and $f_2^w \leq \tilde{f}_2^w$. To show $f_2^b + f_2^w \geq \hat{f}_2^b + \tilde{f}_2^w$, first note that this holds trivially for Cases 1-a, 1-b, 1-d, 2-b and 2-d. In Subarea 3-b, note that

$$f_2^b + f_2^w - \tilde{f}_2^b - \hat{f}_2^w = (f_2^b - \hat{f}_2^b) + (f_2^w - \tilde{f}_2^w) = \frac{s_2}{s_1 + s_2} (N_w - N') - \frac{s_2}{s_1 + s_2} (N_w - N'') = \frac{s_2}{s_1 + s_2} (N'' - N') > 0$$

For Subarea 3-c,

$$f_2^b + f_2^w - \tilde{f}_2^b - \hat{f}_2^w = \frac{1}{s_1 + s_2} (s_1N'_b - s_2N'_w + s_2N'') = \frac{s_2}{s_1 + s_2} (N'' - N') > 0$$

The reader can verify the same for 3-d and 3-e. We proceed to examine the change of cost from NTE to SO.

### Table 2

<table>
<thead>
<tr>
<th>Subarea</th>
<th>Flow change on route 2</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group $w, f_2^w$</td>
<td>Group $b, f_2^b$</td>
</tr>
<tr>
<td>1-a</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1-b</td>
<td>0</td>
<td>+ by $f_2^b$ in (51)</td>
</tr>
<tr>
<td>1-d</td>
<td>0</td>
<td>+ by $N_b$</td>
</tr>
<tr>
<td>2-b</td>
<td>0</td>
<td>+ by $s_2N_b (1 - \frac{s_2}{s_1})$</td>
</tr>
<tr>
<td>2-d</td>
<td>0</td>
<td>+ by $f_2^b$ in (33)</td>
</tr>
<tr>
<td>3-b</td>
<td>– by $f_2^w$ in (36)</td>
<td>+ by $\frac{s_2}{s_1+s_2} (N_w - N')$</td>
</tr>
<tr>
<td>3-c</td>
<td>– by $\frac{s_2}{s_1+s_2} (N_w - N'')$</td>
<td>+ by $\frac{N_b}{s_1}$</td>
</tr>
<tr>
<td>3-d</td>
<td>– by $f_2^b$ in (37)</td>
<td>+ by $f_2^b$ in (37)</td>
</tr>
<tr>
<td>3-e</td>
<td>– by $\frac{s_2}{s_1+s_2} N_w$</td>
<td>+ by $f_2^b$ in (37)</td>
</tr>
</tbody>
</table>

![Fig. 7. An illustration of nine NTE-SO transition subareas.](image)
Subarea 1-a: route choice is not changed, i.e., both groups only use the fast route but switch the order of departure time. As a consequence of the switch, group b is worse off and group w is better off, as analyzed in Table 1.

Subareas 1-b and 1-d: no group b user is on route 2 at NTE, which suggests that their commute at NTE must be smaller than \( z_a T_w \). However, the group b users who select route 2 at SO must bear a higher or equal cost than \( z_a T_b \). Thus all group b users’ cost must increase at SO in these cases, because all group b users must have the same commute cost. Since group w users always stay on route 1 and enjoy the congestion relief, they are clearly better off (note that they are better off even without considering congestion relief, cf. Table 1).

Subareas 2-d and 3-d: route 2 is only used by group b at SO. Because the total flow on route 2 increases at SO (contributed solely by group b), the queue would be longer and the commute cost would be higher on that route if no toll is implemented. While the SO toll actually eliminates the queue, it does so while keeping everyone on the route break even (note that route 2 is homogenous with group b). Thus, the cost at SO increases on route 2, which means that all group b users’ cost increases. On the other hand, group w is the only group that uses route 1 at SO, and thus would benefit from the congestion relief on route 1 due to the net flow reduction.

Subareas 2-b and 3-b: the increase of group b’s cost can be shown similarly as in 2-d and 3-d. Note that route 1 is now shared by both groups at SO. To see why the cost of group w will be lower, note that route 1 is less congested at SO, and that the SO toll would further reduce the cost of group w on route 1.

Subarea 3-e: both groups b and w would use route 2 at SO. Again, the net flow increase on route 2 would lead to longer queue and higher commute cost when there is no toll. The introduction of SO toll will actually eliminate the queue, yet further increase the cost of group b. Consequently, the commute cost of group b user will increase at SO. The cost of group w will be reduced because route 1, which is only used by group w, is now less congested.

Subarea 3-c: the increase of group b’s cost can be established similarly as in 3-e. The reduction of group w’s cost can be shown using the similar argument used in subareas 2-b and 3-b.

This completes the proof. □

Table 2 reveals that the usage of the slow route (route 2) is always greater at the system optimum. This finding agrees with the results found using the static model with two competitive routes (see e.g. Verhoef et al., 1996). Moreover, Theorem 3 highlights the well-known inequitable welfare effects of the SO toll: it favors users with higher VOT and more flexible work schedule and penalizes those with lower VOT and more rigid schedule.

As suggested in the literature (Small, 1992; Adler and Cetin, 2001; Eliasson, 2001; Kockelman and Kalmanje, 2005; Nie and Liu, 2010), the equity issue may be addressed through a lump-sum refunding of toll revenues. Note that the total revenue from the marginal SO toll scheme can be computed as the area under the toll curve as shown in Fig. 2(b). In the two-route case, the total revenue \( TR \) takes the general form

\[
TR = \sum_{j=1,2} \frac{1}{2S_j} \left( \delta a f_j^s (f_j^b + f_j^w) + (\delta a f_j^b + \delta w f_j^b) f_j^w \right)
\]

If TR is refunded equally to all road users, each will receive \( TR/(N_a + N_b) \). Consequently, the condition that guarantees Pareto-improving (i.e., nobody’s commute cost increases) after pricing and refunding can be found for each subarea. For instance, without refunding, the costs at SO and NTE in Subarea 3-b can be computed by the marginal cost in Eqs. (44) and (27) and (28), respectively. Only group b needs concern us here because Theorem 3 indicates that its cost will rise after pricing. Recalling Eqs. (35)-(38) and 50 and 51, the increase of commute cost can be formulated as follows:

\[
\Delta c_b = mc_b - c_b = \frac{N_w}{s_1 + s_2} \left( \delta_b - \delta_w \frac{z_b}{z_w} \right) > 0
\]

The inequality follows from (3) and Assumption (1). To ensure Pareto-improving after lump-sum refunding thus requires

\[
\frac{N_w}{s_1 + s_2} (N_b + N_w) \left( \delta_b - \delta_w \frac{z_b}{z_w} \right) \leq TR
\]

While (64) can be further reduced to gain more insights, we leave this laborious work to future studies. Suffice it to mention here that (64) is not always satisfied, as to be shown in our numerical studies.

4.3. Time-based system optimum

As mentioned in the single-route problem, each group’s schedule delay is valued by \( \beta_i/s_i \) when system cost is measured in time. Therefore, group b would travel at the center of the rush hour because it has larger \( \beta_i/s_i \) ratio by Assumption (1). Accordingly, finding TSO flow pattern can be formulated as the following optimization problem (using the result from the single-route problem):

\[
\min \sum_{j=1,2} \frac{1}{2S_j} \left( \delta a f_j^s (f_j^b + f_j^w) + (\delta a f_j^b + \delta w f_j^b) f_j^w \right)
\]
examine the impacts of the marginal SO toll on route choice and welfare when Assumption (1) and (2) hold.

5. Numerical studies

Parameters adopted in the numerical studies.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Travel time (min)</th>
<th>VOT parameters ($/h)$</th>
<th>Capacity (veh/h)</th>
<th>Time (N/A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category</td>
<td>$T_1$</td>
<td>$T_2$</td>
<td>$a_w$</td>
<td>$a_b$</td>
</tr>
<tr>
<td>Value</td>
<td>18 min</td>
<td>30 min</td>
<td>6.4</td>
<td>2.5</td>
</tr>
</tbody>
</table>

The reader can verify that the Hessian of $TT(f)$ is positive definite, and therefore the above optimization problem is strictly convex and has a unique minimum.

**Theorem 4.** The class-specific route flows are the same at NTE and TSO.

**Proof.** Multiplying $a_w$ and $a_b$ in (68) and (69) respectively yields

\[ f_w(t_w) = T_j + \frac{1}{s_j} \left( \frac{\delta w f_w}{a_w} + \frac{\delta w f_w}{a_w} \right), \]

\[ f_b(t_b) = T_j + \frac{1}{s_j} \left( \frac{\delta w f_w}{a_w} + \frac{\delta w f_w}{a_w} \right), \]

where $\lambda_w$ and $\lambda_b$ are multipliers associated with demand constraint (65), and

\[ mt_w = T_j + \frac{1}{s_j} \left( \frac{\delta w f_w}{a_w} + \frac{\delta w f_w}{a_w} \right). \]

\[ mt_b = T_j + \frac{1}{s_j} \left( \frac{\delta w f_w}{a_w} + \frac{\delta w f_w}{a_w} \right). \]

The complementarity conditions in KKT read:

\[ f_w(t_w) = 0, \quad \lambda_w \geq 0, \quad \text{for } j = 1, 2 \]

\[ f_b(t_b) = 0, \quad \lambda_b \geq 0, \quad \text{for } j = 1, 2 \]

which is equivalent to the equilibrium condition at NTE. Consequently, the flow distribution patterns will be identical at TSO and NTE.

Note that all travelers break even when the marginal TSO toll is implemented, because neither departure order nor route choice is changed from NTE to TSO. Thus, the marginal TSO toll is Pareto-improving. However, Section 3.3 shows that the marginal TSO toll is generally discriminatory, and hence is difficult to implement. While it is possible to find an alternative TSO toll that is anonymous (see Fig. 4(b)), such a toll may introduce extra costs and therefore may not be Pareto-improving.

5. Numerical studies

In this section, numerical studies are conducted to verify the analytical results obtained in the previous sections. We first examine the impacts of the marginal SO toll on route choice and welfare when Assumption (1) and (2) hold. Then, the flow patterns at SO and TSO are compared, as well as their welfare effects. Lump-sum revenue refunding schemes for the SO case are also tested. Such schemes may or may not fully overcome the inequity issue, as demonstrated in two separate examples. Finally, a sensitivity analysis is used to examine how route usage varies with the VOT parameters. Unless otherwise specified, the parameters listed in Table 3 are adopted throughout this section.

5.1. Route usage and welfare effects of SO toll

We first solve NTE and SO problems for different demand patterns ($N_w, N_b$), where $N_w$ and $N_b$ vary within the range [0, 1500] (veh). The contour plots in Fig. 8 visualizes how group-specific flow distributions change after the marginal SO toll.

![Figure 8](image-url)
is imposed. The reported flow changes are constructed for the slow route, i.e., route 2, and the horizontal and vertical axes in the plot indicate the demands of groups b (blue-collar) and w (white-collar), respectively. Also note that a cell with lighter color in the contour plots always represents a larger flow "increase" on route 2. Thus, the dark top area in Fig. 8(a) indicates that the white-collar workers travel less on route 2 at SO than NTE. The bottom area in the same plot is white, indicating no changes in route choice for the white-collar workers from NTE and SO. Note that the two areas are distinguished by a boundary line $N_w = N'' = 990$ veh ($N''$ is defined in Eq. (29)). Similarly, Fig. 8(b) shows that the blue-collar workers generally use the slower route more at SO than NTE, and their route choice remain unaffected only in the black lower-left area, which is corresponding to low demand levels.

Fig. 8(c) shows that the total flow on route 2 always increases from NTE to SO. That is, some users are “totted off” the fast route to achieve the desirable congestion relief for the system as a whole. Moreover, all totted-off users are blue-collar workers as observed from Fig. 8(a) and (b). Thus, the SO toll maximizes the social welfare at the expense of the blue-collar workers. We note that this result verifies Theorem 3. Interestingly, Fig. 8(c) also suggests that the SO-induced flow surge on the route 2 does not keep increasing with the demand level. Instead it has an upper bound, which is achieved in Subareas 3-b and 3-c (about 85 as revealed in the figure). As shown in the proof of Theorem 3, the total flow change in both subarea 3-b and 3-c equals $s_2(N'' - N')/(s_1 + s_2)$, which depends only on the capacity and free flow travel time on both routes (not the level of demands).

Contour plots presented in Fig. 9 compare individual user’s commute cost at NTE and SO. A cell with lighter color in the plots represents a larger positive cost increase. As shown in Fig. 9(a), the commute cost of group b always increases at SO, which indicates that group b is worse off at all demand levels in this setting. In contrast, group w benefits from the SO toll since their costs are always lower at SO than at NTE (see Fig. 9(b)). Note that the contour plots clearly delineates the boundary lines as analyzed in Fig. 7. The reader can verify that here $N' = 773, N'' = 990, N'_w = 289, N''_w = 1206$. It is worth noting that the benefit of group w generally increases with $N_w$. That is, for the same $N_w$, the larger is the $N_b$, the more does the group...
w benefit from the SO toll. The opposite is true for the group b in most case (except in the relatively small subareas 1-d, 2-d and 3-d): for the same \( N_w \), the larger is the \( N_w \), the more does the group b suffer from the SO toll. This observation highlights the unequitable position of the two groups in relation to the SO toll.

5.2. SO vs. TSO

To compare the SO and TSO toll schemes, a fixed demand pattern \( N_w = 1100, N_b = 1900 \) is used. Note that the given flow pattern corresponds to NTE-SO subarea 3-b in Fig. 7. At NTE, both routes are used by groups b and w; at SO, all group w users are assigned to the fast route, and group b users are distributed among both routes (with more flow on the slow route than at NTE); at TSO, the route usage remains the same as at NTE. The route choice patterns reported in Table 4 confirm the above analysis.

For more detailed analysis, the departure and arrival patterns of both groups on route 1 are plotted in Fig. 10. From Fig. 10(a) we observe that (1) the two groups had to switch the departure order to achieve SO (note that group b travels at the center at NTE but switch to tails at SO), and (2) the rush hour starts 2.1 min later and ends 0.5 min earlier at SO. The latter is evidently due to the reduction of the total flow on route 1, which help alleviate congestion. Fig. 10(b) demonstrate that TSO retains the same departure order and arrival curve as at NTE. The total system times and costs at NTE, SO and TSO are compared in Table 5. As expected, the minimum system time is 1707.3 h, achieved at TSO, and the minimum system monetary value (without toll) is $5812.2, achieved at SO. However, the total travel time at SO is only 3.9% higher than the optimal value, whereas the total cost at TSO is more than 20% higher than the minimum value. From the perspective of overall welfare improvement, the SO toll is more desirable than TSO as it significantly improves the monetary value (more than 35%) and yet still provides near-optimal solution when the cost is measured in the unit of travel time. However, a TSO toll is attractive in that it is Pareto-improving, that is, no user is worse off after pricing. Moreover, although the total system monetary value of TSO is larger than that of SO, it still represents more than 20% reduction from the NTE case. As a consequence, TSO provides an alternative toll scheme that is more equitable in the presence of user heterogeneity.

We proceed to examine the possibility of using revenue redistribution to overcome inequity created by the SO toll. The individual cost of each group at NTE and SO are reported in Table 6, where SO is achieved by implementing the marginal SO toll specified in Eq. (13) on both routes. The total revenue from the toll can be computed from Eq. (63) as $2214.61. If we refund this revenue equally to all travelers as a lump-sum return (that is, each traveler receives $2214.61/3000 = $0.74), the individual costs will be reduced. As shown in the fourth column in Table 6, this simple lump-sum refunding scheme fully compensates group b’s loss, which makes the SO toll Pareto-improving.

### Table 4

Route usages in NTE and SO.

<table>
<thead>
<tr>
<th></th>
<th>( f_w^1 )</th>
<th>( f_w^2 )</th>
<th>( f_b^1 )</th>
<th>( f_b^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTE</td>
<td>1056</td>
<td>44</td>
<td>1140</td>
<td>760</td>
</tr>
<tr>
<td>SO</td>
<td>1100</td>
<td>0</td>
<td>1009</td>
<td>891</td>
</tr>
<tr>
<td>TSO</td>
<td>1056</td>
<td>44</td>
<td>1140</td>
<td>760</td>
</tr>
</tbody>
</table>

Fig. 9. Comparison of user’s cost at NTE and SO.
Table 5
Total travel times and monetary value at NTE, SO and TSO.

<table>
<thead>
<tr>
<th></th>
<th>Total time (h)</th>
<th>Total cost without toll ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTE</td>
<td>2353.8</td>
<td>9075.3</td>
</tr>
<tr>
<td>TSO</td>
<td>1707.3</td>
<td>7018.6</td>
</tr>
<tr>
<td>SO</td>
<td>1773.3</td>
<td>5812.2</td>
</tr>
</tbody>
</table>

Fig. 10. Departure and arrival patterns on route 1 at NTE, SO and TSO.
Moreover, note that the Pareto-improving can be achieved without returning all toll revenue. In this example, a simple calculation indicates that, as long as the refund is larger than \((2.11 - 2.02) \times 3000 = \$270\), Pareto-improving is always guaranteed. Thus, the pricing–refunding scheme provides an opportunity to simultaneously (1) improve total social welfare, (2) produce revenue that can be spent on infrastructure and/or alternative modes, and (3) make everyone better off.

It is worth noting that Pareto-improving through lump-sum refunding is not always feasible. For example, for Subarea 3-b, Pareto-improving is ensured only when Condition (64) is met. To provide an example in which lump-sum refunding does not warrant Pareto-improving, a set of parameters listed in Table 7 are employed. Further, to locate this example also in the Subarea 3-b, we set \(N_w = 2200\), \(N_b = 800\). Table 8 reports the individual cost of each group at NTE and SO, with and without refunding. Clearly, refunding is not enough to compensate the loss of group

\[
\text{Group } b: \quad \text{cost} = 2.47 + 4.31 = 6.78 \quad \text{with refunding}
\]

\[
\text{Group } b: \quad \text{cost} = 2.47 + 4.31 = 6.78 \quad \text{no refunding}
\]

5.3. Sensitivity analysis of VOT parameters

An important consequence of user heterogeneity is the different flow distributions at NTE and SO. As proven in Theorem 3, the SO toll will drive more group \(b\) users from route 1 to 2, and accordingly increase the total flow on route 2. The numerical example presented in Section 5.1 shows that the total flow increase on route 2 is bounded from the above and does not keep rising with demand levels. The bound revealed from that example is approximately 85, which is relatively small. It is useful to reveal how this bound would be affected by the VOT parameters, which is pursued in the following sensitivity test.

In this test, we fix all parameters except \(\alpha_b\) and \(\beta_b\). For each combination of \(\alpha_b\) and \(\beta_b\) such that \(\alpha_b/\beta_b < \alpha_w/\beta_w, \beta_b < \beta_w\), the maximum flow increase on route 2 is determined by choosing a pair of demands \((N_w, N_b)\) within the Subareas 3-b or 3-c. In our test, we simply choose \((N_w = N_w^* + 1, N_b = N_b^* + 1)\) where \(N_w^*\) and \(N_b^*\) are defined in Eq. (45). The contour plots in Fig. 11 shows how the maximum flow increase on route 2, denoted as \(\max(\Delta f_2)\), varies with \(\beta_b\) and \(\alpha_b\). As shown in Fig. 11(a), \(\max(\Delta f_2)\) varies from zero to about 392.7 within the feasible region of \(\alpha_b\) and \(\beta_b\), decreasing with \(\alpha_b/\beta_b\). The value in the infeasible region is always set to zero. In fact, \(\max(\Delta f_2)\) can always be obtained analytically. Recalling that \(\max(\Delta f_2)\) is always achieved within Subarea 3-c or 3-b (Fig. 8) and using the result in Table 2, \(\max(\Delta f_2)\) corresponding to Subarea 3-c can be obtained as

\[
\max(\Delta f_2) = \frac{1}{s_1 + s_2} (s_1 N_b^* - s_2 N_w^* + s_2 N_w^*) = \frac{s_1 s_2}{\phi(s_1 + s_2)} \left( \frac{\alpha_w}{\beta_w} - \frac{\alpha_b}{\beta_b} \right) (T_2 - T_1)
\]
Since only $a_b$ and $b_b$ are variables in our experiment, the above formula suggests that $\max(\Delta f_2)$ should decrease linearly with $a_b/b_b$, as depicted in Fig. 11(a). Fig. 11(b) shows the contour plot of $\max(\Delta f_2)$ relative to the total demand $N_b + N_w$. Note that the relative change is no longer a linear function of $a_b/b_b$. Rather, the contour lines bend over when $b_b$ increases, suggesting that larger $b_b$ reduces the relative significance of flow change on route 2. The largest observed percentage change in this setting is 20.8%.

6. Conclusions

In the spirit of Vickrey (1969), this paper extends the bottleneck analysis to study congestion behavior of morning commute and its implications to transportation economics. Our analysis differs from those in the literature in two aspects. First, it considers simultaneous route and departure time choices of heterogenous users who are identified mainly by their valuation of travel time and punctual arrival. Second, it introduces alternative system optima and the corresponding toll schemes: one minimizes system cost in the unit of monetary value (i.e., conventional SO) and the other minimizes it in the unit of travel time (i.e., time-based SO, or TSO). Our main findings are summarized in the following.

- A TSO toll provides a Pareto-improving alternative to the SO toll, which is known to create inequitable welfare effects among heterogenous users. Specifically, a marginal TSO toll does not increase any user's commute costs, whereas an SO toll generally benefits wealthier users at the expense of the poorer ones. However, a Pareto-improving SO toll scheme may be obtained, albeit by no means guaranteed, through refunding toll revenue as a lump-sum return.
- Unlike an SO toll, a TSO toll may not be anonymous in general. That is, users may have to pay tolls according to their VOT parameters, even if they arrive at the bottleneck at the same time. Since imposing such a discriminatory toll may be neither technically feasible (as these VOT parameters are not easy to observe and/or quantify) nor politically acceptable, it is hard to implement a TSO toll in practice. However, an anonymous (non-marginal) TSO toll may exist when certain conditions are met.
- Route choice at no-toll equilibrium is no longer optimal at SO when user heterogeneity is considered. Specifically, the use of the faster (hence more congested) road has to be reduced in order to minimize the system cost. Thus, an SO toll may not only place the wealthier users in a more desirable departure time slot, but also help them travel more on the faster route. It does exactly the opposite to the poorer users. Consequently, the social inequity caused by the SO toll is worsened by the consideration of route choice.

Although the TSO toll provides an appealing pricing alternative, its discriminatory nature makes its implementation a practical challenge. Our analysis suggests that conditions may arise which would enable anonymous TSO tolls (e.g., when $\beta_w < \beta_b$ and Assumption (1) is satisfied). Deriving more general existence conditions for anonymous TSO tolls justifies further investigations. Another possible direction for future research is Pareto-improving refunding schemes for the SO toll schemes, which holds promise in resolving the equity issue.

The analyses presented herein may also be extended to the models with more than two user groups and/or more than two parallel routes, or more general heterogeneity settings (e.g., allowing different users to have different ratios of late to early...
penalty costs). Arnott et al. (1994) discussed the case of three user groups in the single-bottleneck problem. Extending their results to the two-route case is relatively straightforward for the no-toll equilibrium (NTE) problem. Specifically, as the demands increase, the blue-collar group would first switch to the slow route, followed by the middle group (the group whose welfare parameters range between those of the blue-collar and white-collar groups), and finally the white-collar group. Solving the SO problem for the three-group-two-route model, however, is expected to be much more complicated in that many more sub-cases could emerge as a result of interactions of the three groups. Visualizing these results would also become considerably more difficult, because it may require a 3-D representation. Similarly, a two-group-three-route model is also much easier to solve for the NTE problem than for the SO problem. While it may be possible that some structure of the solution can be utilized to streamline the presentation, the tractability of such analyses, as well as the usefulness of the insights generated by them, degrades quickly as the complexity of the model increases. Numerical methods, such as those based on DTA models, may be a better tool if the solution to a more realistic model is desired. Discussions of those methods, however, are beyond the scope of this paper.

Acknowledgements

We would like to thank two anonymous referees for their constructive comments on an earlier version of this paper. In particular, the parallels between Theorem 1 and some results from Arnott et al. (1992) were initially missed in our literature review and were brought up by one of the referees. We accept full responsibility for remaining shortcomings and errors.

Appendix A. List of notation

<table>
<thead>
<tr>
<th>Supply parameter</th>
<th>( T_i ) / ( s_i )</th>
<th>free flow travel time on route ( j ), ( T_1 &lt; T_2 ) capacity of route ( j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demand parameter</td>
<td>( t^* )</td>
<td>identical preferred arrival time across users</td>
</tr>
<tr>
<td></td>
<td>( \alpha_i )</td>
<td>unit cost of travel time of group ( i )</td>
</tr>
<tr>
<td></td>
<td>( \beta_i )</td>
<td>unit cost of early arrival of group ( i )</td>
</tr>
<tr>
<td></td>
<td>( \gamma_i )</td>
<td>unit cost of late arrival of group ( i )</td>
</tr>
<tr>
<td></td>
<td>( \beta_i &lt; \alpha_i &lt; \gamma_i )</td>
<td></td>
</tr>
<tr>
<td>VOT related</td>
<td>( \eta )</td>
<td>( \gamma_i / \beta_i ), relative cost of late to early arrival</td>
</tr>
<tr>
<td></td>
<td>( \phi )</td>
<td>( \eta / (1 + \eta) )</td>
</tr>
<tr>
<td></td>
<td>( \theta_i )</td>
<td>( a_i / (\alpha_i - \beta_i) )</td>
</tr>
<tr>
<td></td>
<td>( \rho_i )</td>
<td>( a_i / (\alpha_i + \gamma_i) )</td>
</tr>
<tr>
<td></td>
<td>( \delta_i )</td>
<td>( \beta_i \gamma_i / (\beta_i + \gamma_i) )</td>
</tr>
<tr>
<td>Demand related</td>
<td>( N_i )</td>
<td>demand of group ( i )</td>
</tr>
<tr>
<td>VOT related</td>
<td>( N^*, N^{**} )</td>
<td>defined in Equation (29)</td>
</tr>
<tr>
<td></td>
<td>( N_{G,i}^<em>, N_{B,i}^</em> )</td>
<td>defined in Equation (45)</td>
</tr>
<tr>
<td>Solutions at NTE&amp;SO</td>
<td>( f_i^j )</td>
<td>flow of class ( i ) on route ( j )</td>
</tr>
<tr>
<td>User cost</td>
<td>( \tau(t) )</td>
<td>marginal toll of commuter departing at time ( t )</td>
</tr>
<tr>
<td></td>
<td>( g(t) )</td>
<td>schedule cost of commuter departing at time ( t )</td>
</tr>
<tr>
<td></td>
<td>( c_i^j )</td>
<td>cost of group ( i ) on route ( j )</td>
</tr>
<tr>
<td></td>
<td>( m_{C,i}^j )</td>
<td>SO marginal cost with respect to ( f_i^j )</td>
</tr>
<tr>
<td></td>
<td>( m_{T,i}^j )</td>
<td>TSO marginal cost with respect to ( f_i^j )</td>
</tr>
<tr>
<td>System performance</td>
<td>( C )</td>
<td>system cost in monetary unit</td>
</tr>
<tr>
<td></td>
<td>( TT )</td>
<td>system cost in travel time unit</td>
</tr>
<tr>
<td></td>
<td>( TR )</td>
<td>toll revenue</td>
</tr>
</tbody>
</table>

Note: when a variable has both subscript and superscript, the subscript \( i \) refers to group \((i = b, w)\) and the superscript \( j \) refers to route \((j = 1, 2)\). Otherwise, subscript may be used to refer either group (such as VOT parameters) or route (such as supply parameters).
Appendix B

This appendix shows how UE costs given in (4) and (5) are obtained. See Fig. 1 for a reference of $t_0 - t_4$. From the standard bottleneck analysis, we know that

$$ t_0 = t^* - \frac{(N_w + N_b) y_w}{s(\beta_w + \gamma_w)} = t^* - \frac{\phi(N_w + N_b)}{s}, \quad t_5 = t^* - \frac{\beta_w}{x_w} \frac{\phi(N_w + N_b)}{s} $$

Let $x = t_5 - t_4$, $y = t_4 - t_5$, and note the following identities.

$$ y \beta_s = (x + y) \beta_b s = \frac{\gamma_b N_b}{\beta_b} = \phi N_b $$

$x$ and $y$ can be solved from the above as:

$$ y = \phi N_b \beta_s, \quad x = \left( \frac{1}{\beta_w} - \frac{1}{\beta_b} \right) \frac{\phi N_b}{s} $$

Thus,

$$ t^* - t_4 = x + t^* - t_5 = \left( \frac{1}{\beta_w} - \frac{1}{\beta_b} \right) \frac{\phi N_b}{s} + \left( 1 - \frac{1}{\beta_w} \right) \frac{\phi(N_w + N_b)}{s} = \left( 1 - \frac{1}{\beta_w} \right) \frac{\phi N_w}{s} + \left( 1 - \frac{1}{\beta_b} \right) \frac{\phi N_b}{s} $$

Since this is the total delay of the person from group $b$ who arrives on time, it gives the commute cost for group $b$ as (using (3))

$$ c_b = \gamma_b (t^* - t_4) = \frac{\lambda_h \beta_w}{\lambda_w} \frac{\phi N_w}{s} + \frac{\phi N_b}{s} = \frac{\lambda_h}{\lambda_w} \delta_s N_w + \frac{\delta_b N_b}{s} $$

The commute cost for group $w$ is

$$ c_w = \frac{\delta_w (N_w + N_b)}{s} $$

References

