Transit priority lanes in the congested road networks

Note: Revisons-4 are highlighted for convenience

Saeed Asadi Bagloee (corresponding author)
PhD candidate
Department of Infrastructure Engineering, Melbourne School of Engineering
The University of Melbourne, Victoria 3010, Australia
Tel: +61-3-90353864 : Email: saeed.bagloee@unimelb.edu.au

Prof Majid Sarvi
Chair in Transport Engineering and the Professor in Transport for Smart Cities
Department of Infrastructure Engineering, Melbourne School of Engineering
The University of Melbourne, Victoria 3010, Australia
Tel: +61-3-83441759, Email: majid.sarvi@unimelb.edu.au

Prof Avishai (Avi) Ceder
Chair in Transportation; Founder and Former Director of Transportation Research Centre
Transportation Research Centre, Department of Civil and Environmental Engineering,
The University of Auckland, 20 Symonds Street, New Zealand
Tel: +972-50-5216084, E-mail: a.ceder@auckland.ac.nz

Abstract

Given the advances in communication technologies and real-time traffic management, transit priority lanes are emerging as an indispensable component of intelligent transport systems. This scheme calls for giving priority to public transport. In this study, the question of interest is: Which roads can be nominated to give an exclusive lane to transit modes? Due to computational and theoretical complexities, the literature has yet to address this problem comprehensively at the network level considering various modes (public and private). Additionally, taking space away from private modes in favor of public transport may adversely affect the congestion level. To this end, inspired by the Braess Paradox, we seek mis-utilized space used by private modes to be dedicated to transit modes mainly on congested roads. To find such candidate roads, we define a merit index based on transit ridership and congestion level. The problem then becomes - to find the best subset of these candidate roads to cede a lane to transit mode. It is formulated as a bilevel mixed-integer, nonlinear programming problem in which the decision variables are binary (1: to cause the respective road to have an exclusive transit lane or 0: not). The adverse effects are minimized on the upper level represented by total travel time (public and private modes) spent on the network. The lower level accounts for a bimodal traffic assignment, to consider the impact of transit priority on private modes. We then develop an efficient low-RAM-intensity branch and bound as a solution algorithm. The search for the subset is made in such a way that improved public transport is achieved at zero cost to the overall performance of the network. A real dataset from the city of Winnipeg, Canada is used for numerical evaluations.

Keywords: Transit priority lane, Braess Paradox, Bilevel, Branch and bound

1. Introduction

Traffic congestion is a chronic challenge in cities across the globe. Congestion pricing and public transport termed *transit* in this paper, have long been viewed as an effective means of addressing the growing travel demand (Ceder, 2015). Given factors such as cost efficiency, environmental concerns, equity and public support; promoting transit even at the cost of a private mode is gaining momentum. The priority of mobility has to be given to people rather than vehicles. Hence, providing priority to transit mode in terms of road space (Mesbah et al., 2011b) and signal time is of interest to practitioners as well as academia (Bagherian et al., 2015; Diab and El-Geneidy, 2013; Guler and Menendez, 2015; Mirchandani et al., 2010). In this study, we aim to address optimal reallocation of road space to transit mode on an existing urban transport network. In particular, we are interested in finding a network of exclusive transit priority lanes in the heart of cities. The concept of priority lanes has been introduced in many cities (Smith and Hensher, 1998). It also has been adopted in special traffic access plans for large scale crowd gathering events such as sports games, concerts, New Year's events etc. (Cova and Johnson, 2003). A good number of studies have investigated the impact of priority initiatives extensively, while little is devoted to the strategic design of such schemes (Bagloee et al., 2016b). For evaluation, a variety of methods based on statistical analysis and simulations have been employed (Eichler and Daganzo, 2006; Li and Ju, 2009; Liu et al., 2006; Sakamoto et al., 2007; Tse et al., 2014; Viegas and Lu, 2004).

As for the strategic design, the subject of transit priority, or in general, the subject of transit network design/planning can be viewed as a general network design problem (NDP), which has been and continues to be, challenging to scholars across many disciplines. For such problems, given a set of candidate projects where each is associated with a cost, the best subset of them subject to a limited budget is sought. The most challenging variant of the NDP is the discrete network design problem (DNDP) in which the decision variables are binary (1: build and 0: not to build the network component).

A typical road can be considered as a directional link consisting of one or more lanes. In general, roads and lanes can be used by all traffic modes (cars and buses). If a road is set to be a transit priority lane, at least one of its lanes is dedicated to exclusive use by buses. Turning a lane into a transit priority lane is also associated with expenses derived from road marking, special signage/signals, lighting, etc. Hence, the transit priority lanes design problem (TPLDP) can be articulated as follows: Given a set of candidate roads, which lanes should be selected as priority lanes while accounting for a limited budget? Taking space away from private modes in favor of public transport is a delicate job, which may adversely lead to gridlock congestion. Therefore, one needs to minimize such adverse impact. One intuitive way is to minimize the total travel time spent in the network which is set as the objective function in the TPLDP. The consensus in the literature is to model the DNDP and the TPLDP as a bilevel mixed-integer, nonlinear optimization problem. The objective function is placed in the upper level, while the lower level accounts for the user equilibrium (UE) private traffic flow as well as the transit flow. It is proven that any bilevel programming problem is NP-hard so that the problem becomes quickly intractable as the size of the problem increases (Ben-Ayed and Blair, 1990).

Among the variety of methods available in the literature, some try to reach an optimum solution but cannot scale to handle large-sized networks, others aim to address large-sized networks at the cost of compromising the quality of the solution.

Irrespective of whether a method is able to handle large networks or not, some require additional software to function as optimization solver engines. Therefore, enriching the existing commercial transport planning software with an easy-to-use module for solving such problems in an affordable computation time is a highly demanding task. To this end, we develop a RAM-efficient branch and bound (BB) method to address the TPLDP tailored to large-sized networks. First, a set of roads with significant transit volume is identified as a candidate set. Second, using a BB method, the possibility of selecting a subset of the candidate roads is investigated, such that the overall performance of the traffic system, including private and transit flows, is not negatively impacted. It is even possible to improve the overall performance of the transport network, as is well illustrated by the Braess Paradox (BP). The BP refers to the fact that adding a new capacity to a transportation network might adversely degrade the traffic circulation. Empirical evidence, as well as mathematical theories, have shown that the presence of the BP is prevalent in real transport networks (Nagurney, 2010); that is, if there exist some closed roads the traffic circulation can be improved. The main idea of this study is to lay out the transit priority lanes on Braess-tainted roads. It is a utilitarian approach to take advantage of this stigma.

Further, we solve the TPLDP while considering the bimodal feature of the traffic flow which enhances the authenticity of the model. The real dataset for Winnipeg, Canada is used to demonstrating the numerical impact of the solutions calculated by the method proposed here.

It is important to note that the structure of a transit system such as stop positions, transit routes, and fleet size remains intact in this demonstration. Changes may occur with respect to some segments of the existing transit routes currently sharing road space (lanes) with private modes, which may come to be dedicated as exclusive lanes. Hence, the challenge is to find these segments without detriment to the traffic circulation.

This paper contributes to the literature on three fronts: (i) A network-wide approach to the TPLDP tailored to large-sized networks of congested roads is developed. (ii) The BP is utilized to nullify the adverse effect of transit priority lanes on the private mode by providing faster public and even private transport. (iii) A RAM-efficient BB algorithm tailored to a bi-modal traffic model is proposed so that its simple structure can easily be embedded in any programming language.

The contents of the remaining sections of the article are organized as specified hereafter. Section 2 presents a review of the literature pertaining to transit priorities. The problem is mathematically formulated in Section 3. The methodology is elaborated in Section 4. Section 5 is dedicated to the numerical evaluations followed by the conclusion in Section 6.

2. Literature review

Due to the computational complexity involved in the TPLDP, a majority of the past studies have taken a localized approach on a specific road or region. Mesbah et al. (2011b) were among the first to address the TPLDP at the network level. They developed algorithms based on the Benders decomposition and genetic algorithm (GA) and applied them to a grid-structured network consisting of 38 nodes and 49 links. Since then the literature is still leaning toward a more localized focus on the subject of transit priority. Basso et al. (2011) studied congestion management policies through numerical analysis of a local pilot area centered around a short-length road and drew conclusions in favor of dedicated bus lanes as the best ad hoc policy. Guler and Cassidy (2012) investigated traffic operations of the exclusive transit lanes at traffic bottlenecks occurring at intersections. They suggested some operational measures for sharing space at the bottlenecks temporarily. Xie et al. (2012) studied intermittent priority of the bus lanes using simulation methods. Zheng and Geroliminis (2013) proposed a macroscopic method of solving the problem of road space allocation based on a fundamental diagram, applicable to the cases of an aggregate size area such as districts. Geroliminis et al. (2014) studied the dynamism of a bimodal traffic flow (public and private modes) using the notion of fundamental diagrams which can then be utilized for traffic management schemes including transit priority.

Yao et al. (2012) developed a bilevel programming method similar to that of Mesbah et al. (Mesbah et al., 2011a) to address optimization of exclusive bus lanes at the network level. The GA has been used as a solution algorithm and the arrivals' headways of the buses were optimized for a network consisting of 13 nodes, 19 links. At the lower level of the formulation, a path-based traffic assignment using the method of successive averages was employed. Hadas and Ceder (2014) addressed the TPLDP at network level through a single level optimization formulation while approximating the impact of congestion. Khoo et al. (2014) also adopted bilevel programming , while a GA was employed as a solution algorithm. For the lower (traffic assignment) problem a simulation software was used at the cost of higher computational time. (Chen, 2015; Sun et al., 2014) also employ the GA to solve a problem similar to that articulated by (Mesbah et al., 2011b). In the work of (Sun et al., 2014) the concerns for the transit enterprise in operating the transit fleet efficiently as a business has also been taken into consideration. The algorithm was applied to a network consisting of 29 origin-destination (OD) pairs and 52 nodes. Wang et al. (2016) also modeled the exclusive bus lane design as a bilevel problem for which they proposed a heuristic method to find the priority lanes as well as the transit assignment for a network of 24 nodes; 76 links. Zhang et al. (2014) were among the first who addressed the simultaneous design of road expansion and transit system. Their formulation is able also to find appropriate transit lanes. The problem was transformed to a single level mixed integer problem. Given the number of binary variables involved in the formulation, application to large-sized networks is open for further investigation.

Bagloee et al. (2016b) present a methodology to find out the priority lanes for bicycles; it is formulated as a multiclass traffic pattern. Moreover, the bicycles are considered as a distinct class based on the Spiess' bias term as a proxy for multiclass traffic while no consideration is made to the public transport.

The adverse impact of the transit priority lane on the private mode has also been the subject of research (Fang et al., 2014; Wu et al., 2013; Wu et al., 2015; Wu et al., 2014; Yingfeng and NaiQi, 2010; YunFei et al., 2011). In this respect, the key point is obviously, to take the interaction of the public and private modes into account which is a nontrivial task. In the above-mentioned works, the adverse impact of the transit priority lanes on private modes is considered a priori (i.e. a given exogenous input). There is also room for improvement to consider the public and private travel demands. Nevertheless, the formulations and the models developed in the above-mentioned articles can be used extensively in uncongested networks. One of the applications could be finding reserved lanes for transport of hazardous material on the outskirts of cities which are usually uncongested space where traffic congestion is not a source of concern (Zhou et al., 2014; Zhou et al., 2012). In a similar fashion, Fang et al. (2013) sought applications of the priority lane for freight transport in which the freight demand was considered exclusively.

Recently Yao et al. (2015) have taken the uncertainty in the travel times and the capacity of the roads as well as travelers' risk adverse behavior into account. Their results underscore the importance of incorporating a combinatorial optimization of the exclusive bus lanes.

The concept of exclusive bus lanes has also been introduced at signalized intersections along with an exclusive bus phase (in the timing of the signals) to reduce the delay of the buses. A recent review on this subject is provided by (Guler et al., 2016).

Network design has recently witnessed a renewed interest in methods of more analytical strength (Fontaine and Minner, 2014; Gao et al., 2005; Wang et al., 2013; Zhang et al., 2014). Such new trends can be partly motivated by the great interest in better deterministic and not the stochastic solution. Furthermore, the recent rise in the use of such methods which come at the cost of greater computational times owes much to the recent advances in computational technology and knowledge of optimization in both hardware and software (Achterberg and Wunderling, 2013; Bixby, 2012; D'Ambrosio and Lodi, 2013). It is noteworthy to highlight the fact that problems such as NDP or TPLDP are of strategic nature (due to their pervasive impact), so that, one can afford a greater computation in the scale of hours or even days.

As can be seen practical applications of the proposed methods are rare in the existing literature. To this end, we developed an analytical algorithm (an efficient branch-and-bound) tailored to the large-sized networks while taking into account the interaction of private and public modes via a bimodal traffic model.

3. Transit priority lanes design problem (TPLDP)

In this section, we first set the mathematical definitions of the TPLDP, and we then elaborate on the way we include the bimodal aspect of the traffic flow.

For ease of formulation, we adopt the following convention: we denote a road considered as a candidate by ℓ (with, for instance, three lanes), and we suppose it is replaced with two new roads $\ell' \in \bar{A}$ and $\ell'' \in A$: (i) road $\ell' \in \bar{A}$ with only one lane which is to be either a mixed mode road or an exclusive transit lane or road and (ii) road $\ell'' \in A$ with two lanes for mixed mode use. We then alternatively refer to them as transit lanes or transit roads (and, by doing so, we can alternatively call them transit priority lanes or transit roads). Therefore, we can have: $\bar{A}$: a set of roads currently with mixed modes (transit and private modes) but considered as candidates for exclusive use by transit modes, and the rest of the roads are denoted by A (the candidate road henceforth is simply called "candidate").

Although, the exposition of the BP in the literature supports a complete closure of the BP-tainted roads, we keep the road open to private cars for two reasons: (i) connectivity: in order to preserve the connectivity of the network, the roads that are closed to cede space to buses must have at least two lanes; in the event they become nominated to give away one lane as a transit priority lane, they still will have at least one lane remaining (ii) optimality: it is proven mathematically that a partial closure (like closing a lane for the transit mode) is more likely to result in a better traffic circulation (this concept is highly exploited in the congestion pricing, (Yang and Huang, 2005) through which cars are discouraged to drive by enforcing a “toll” instead of by any physical restriction. It is important to note that any changes to the network or the travel demand such as adding a priority lane may change the BP’s status (Aashtiani and Poorzahedy, 2004; Nagurney, 2010). Nevertheless, in the formulation of the problem (i.e., in the objective function) we ensure that these changes result in a better traffic circulation across the network. Let us define the following notation:

N : set of nodes,

B : budget available to cover the costs of transit lane implementations such as marking, pavement, curb raising, etc.

y_a : binary decision variable associated with candidate $a \in \bar{A}$; 1: to be used as exclusive transit lane and 0: to remain mixed use road or lane,

c_a : implementation cost associated with candidate $a \in \bar{A}$.

$\bar{x}_a$: hourly public passenger volume on road $a \in A \cup \bar{A}$,

$x_a, \bar{\bar{x}}_a$: hourly private and transit traffic flow in hourly passenger car equivalent or unit (“pce” or “pcu”) on road $a \in A \cup \bar{A}$, respectively, (Note, (i) the network available to the private and transit roads/lanes are A and $A \cup \bar{A}$, respectively, hence $x_a, \bar{\bar{x}}_a \geq 0$ for $a \in A$ and $x_a = 0, \bar{\bar{x}}_a \geq 0$ for $a \in \bar{A}$, (ii) $\bar{x}_a$ is the hourly volume of passenger traffic on the road while $\bar{\bar{x}}_a$ is the car equivalent value of the corresponding number of buses on the respective road $a \in A \cup \bar{A}$, (iii) the pce reflects the physical and operational characteristics of the buses that can vary from 1.2 to 4.5. In traffic models, these values have already assigned to the transit fleet. For instance for the Winnipeg traffic model the pce is either 2 or 2.5 depending on the type of bus).

$t_a(x_a + \bar{\bar{x}}_a)$: general travel time of link $a \in A \cup \bar{A}$, a non-decreasing function of the flow $x_a + \bar{\bar{x}}_a$ of the traffic (called the delay function) for which we comply with a widely used function developed by the U.S. Bureau of Public Roads (BPR) (Sheffi, 1985; Spiess, 1990). Note that the background traffic, $\bar{\bar{x}}_a$,is a fixed value . In addition, provided the delay functions for different configurations of a same road (i.e. one-lane, two-lane etc) are available, switching the delay functions between with/without priority lanes is technically a trivial task in our proposed methodology. To this end, generally speaking, the new delay function must be calibrated based on the field survey data. (For example, two one-lane links, when joined together side by side is NOT simply a 2-lane road. Any stalled vehicle in the one-lane link would close it, while this is not true for the 2-lane road. There is a synergy achieved by additional links. Sometimes, widening ascends the status of a link from, say, local road to arterial, etc. which calls for additional efforts to calibrate appropriate delay functions). Nevertheless, we are employing the BPR delay functions of free flow speed and capacity. In our formulation, as the number of lanes change, the capacities are updated accordingly. However, we keep the free flow speed intact for the following reason. The priority lanes are sought among the congested roads (there is no point to give priority to the mass transit in the uncongested roads). Therefore it is conceivable that a congested road (which is not governed by the free flow speed) after giving away a lane is still congested with delays remaining more or less same.

A_n^-, A_n^+ : set of links starting and ending at node n , respectively, $A_n^-, A_n^+ \subset A \cup \bar{A}$,

x_a : hourly traffic volume in auto or private mode,

R : set of OD pairs $R \subset N^2$,

q_r : hourly travel demand in pcu for OD pair $r \in R$ pertaining to the auto mode.

g_{ij} : hourly transit passenger demand from node i to destination node j . In order to simplify the notation, we define $g_j = -\sum_{i \in N-\{j\}} g_{ij}$ that is the total trip attraction to node j , see (Spiess and Florian, 1989),

P_r : set of paths between OD pair $r \in R$ available to the auto mode,

h_k , : hourly traffic flow on paths $k \in P_r$, pertaining to the auto mode,

$\delta_{a,k}$: road-path incident index, 1 if road $a \in A \cup \bar{A}$ belongs to path $k \in P_r$ pertaining to the auto mode, and 0 otherwise

w_n : average waiting time at node $n \in N$ pertaining to transit system,

f_a : sum of frequency of service for all transit lines on road $a \in A \cup \bar{A}$.

The bilevel TPLDP may be written as follows (note, all variables and parameters are considered non-negative unless otherwise stated):

$$\min_{x_a, \bar{x}_a, y_a} \sum_{a \in A \cup \bar{A}} (x_a + \bar{\bar{x}}_a).t_a(x_a + \bar{\bar{x}}_a) \tag{1}$$

$$\text{s.t.} \quad y_a = 1 \text{ or } 0, \qquad a \in \bar{A} \tag{2}$$

$$\sum_{a \in \bar{A}} c_a.y_a \le B \tag{3}$$

$$\arg\min \sum_{a \in A \cup \bar{A}} \int_0^{x_a} t_a(x_a + \bar{\bar{x}}_a)\,dx \tag{4}$$

$$\text{s.t.} \quad \sum_{k \in P_r} h_k = q_r, \qquad r \in R \tag{5}$$

$$x_a = \sum_{r \in R} \sum_{k \in P_r} h_k \delta_{a,k} \qquad \delta_{a,k} = \begin{cases} 1 & a \in k \\ 0 & a \notin k \end{cases}, \quad a \in A \cup \bar{A} \tag{6}$$

$$\bar{\bar{x}}_a \le U \qquad a \in \bar{A} \tag{7}$$

$$x_a \le (1 - y_a).U \qquad a \in \bar{A} \tag{8}$$

$$\left\{ \begin{array}{ll} \bar{x}_a \in \arg\min \displaystyle\sum_{a \in A \cup \bar{A}} \bar{x}_a.\bar{t}_a + \sum_{n \in N} w_n, & \\ subject\,to & \\ \displaystyle\sum_{a \in A_i^+} \bar{x}_a - \sum_{a \in A_i^-} \bar{x}_a = g_i & i \in N \\ \bar{x}_a \le f_a.w_n, & a \in A_n^+, n \in N, \\ \bar{x}_a \ge 0, & a \in A \cup \bar{A} \end{array} \right\} \tag{9}$$

where x_a, $\bar{x}_a$ are the solutions of the problems (4 ..8) and (9), respectively. The objective function (1) describes the upper-level goal of minimizing the total travel time. Constraints (2) and (3) ensure the feasibility of the candidates with respect to the costs and the available budget. At the lower level ((4), (5), (6)), the Beckmann formulation of the UE flow pertaining to the private mode is computed (Beckmann et al., 1956). Constraints (2), (7) and (8) ensure that private flow will not enter the dedicated transit lanes. (U is a sufficiently large value, say the total demand $\sum_r q_r$). Although constraint (7) is redundant, it is placed within the constraints to emphasize that buses can use candidate roads either exclusively (if it turns out to be $y_a = 1$) or mixed with private mode (i.e., $y_a = 0$). If it is decided that candidate a is to be an exclusive transit lane/road (i.e., $y_a = 1$), then constraint (8) ensures the respective road be closed to the private mode (i.e., $1 - y_a = 0$). Sub-problem (9) carries out transit assignment based on an optimal strategy (Spiess and Florian, 1989) and it returns $\bar{x}_a$ as passenger traffic volume per hour. The sub-problem also returns effective frequency of the transit lanes (or the number of buses) on the roads (note that the roads are also associated with transit delay functions which are functions of travel times experienced by the auto mode). The equivalent value of buses in pcu (denoted by $\bar{\bar{x}}_a$) is then considered as background traffic in the traffic assignment (Spiess, 1984).

At the lower level, a combination of traffic and transit assignment (bimodal) theoretically leads to a nonconvex programming problem. Such problems then require some computationally expensive methods such as variational inequality, not to mention some unresolved issues such as uniqueness and stability of the solutions (Florian and Morosan, 2014). The relevant studies either fall short of fully considering the simultaneous interaction between private and public mode or suffer from lengthy computation time (De Cea et al., 2005; Liu and Meng, 2012). Given these complexities, the above formulation ((4)..(9)) is proven to be able to solve the bimodal traffic assignment adequately, so it has been adopted in some commercial planning applications (INRO, 2009). In this study, formulation (4)..(9) has been coded as a module in EMME 3 (INRO, 2009) and it is called upon by the BB algorithm whenever needed.

Given the fact that the main tenet of a transit priority scheme is to give priority to people and not the vehicle, one may prefer to set up the objective function as minimization of the passenger-time and not the pce-time (see Eq (1)). The answer is three-fold: (i) the measuring unit of congestion, perhaps, unfortunately, is a vehicle not a passenger. In other words, it is customary to count a number of vehicles using traffic monitoring devices and it is very difficult to count the number of occupants. Therefore in order to assess any traffic mitigation scenario (including transit priority schemes), vehicle-based survey methods are more preferable (ii), adhering to minimization of the pce-time has an important implication for the operational costs of the transit fleet. Imagine an **empty** road and a transit ridership of 40 passengers which can be carried by one bus or more. In this scenario, the total travel time of this particular road remains the same and constant because it is empty and one or two more vehicles do not change the travel time significantly. In the case of minimization of the passenger-time, the outcome will become indifferent to the number of buses, therefore the operational costs of the public transport (fleet size) are largely overlooked. (iii) in multi-objective optimization such as minimization of the travel time as well as pollutant, unit of measuring the objective functions are mainly vehicles, not passengers. In other words, adhering to minimization of the pce-time leave space to include other objective functions (such as pollutant's minimization) in the above formulation.

4. Methodology

- **Branch and bound in the context of optimization methods**

There are a number of methods proposed to address the bilevel mixed-integer nonlinear optimization problem. In some, all the combinations of the integer decision variables are evaluated and hence the global optimum solution is guaranteed. The most notable method of enumeration is BB which uses a tree structure to process all the combinations. In the minimization problems, as the tree expands, a lower bound is calculated at each node and the branching is frozen (fathomed) wherever the lower bounds are found greater than the best-found solution (which is also an upper bound value and is called incumbent value). It is evident that as the size of the problem (number of decision variables) increases, the method becomes computationally prohibitive.

Many difficult problems are in fact easy problems complicated by a relatively small number of constraints (Fisher, 2004). Therefore, some scholars tend to split the problem into easy and hard parts: (i) tractable which are solvable very easily and (ii) intractable which are difficult to solve. Examples are Lagrangean relaxation, generalized Benders decomposition, outer approximation and Dantzig-Wolfe. Generally, these methods require an optimization solver engine or a series of convoluted matrix computations. Alternatively, the special BB developed in this study can be coded in any application and does not need any special software.

LeBlanc (1975) proposed a BB method for the DNDP, but due to the computational technology available at the time, it was considered inefficient. In this study, we try to customize the structure of the BB to the TPLDP finely in order to achieve a more efficient algorithm. As the result, a new method for constructing the tree structure at node selection and branching based on the concept of merit index (Bagloee et al., 2016b) is proposed.

The ways the structure of the tree is formed, as well as the presence of the merit index, have a significant impact on the efficiency of the BB algorithm; it results in a less RAM intensive and memoryless algorithm. In the sequel, we discuss how to initialize the BB based on the merit index and how to arrive at a tighter lower bound value to serve the purpose of making the BB as efficient as possible. A detailed discussion on the lower bound values, node selection and branching rules, as well as the termination conditions, are provided in Appendix A.

- **Merit index to find candidate roads**

The first stage is to come up with a set of candidate roads. Looking for transit priority lanes in a suburb or uncongested roads bears no fruit. Accordingly, in this study, we assume the challenging task of laying down the transit priority lane network in the congested parts of the urban road network – namely downtown or CBDs.

Based on the concept of the BP, we look for some roads – though congested – to take away one lane for public transport, without worsening the current congestion. In doing so, we define a merit index based on which the roads are sorted in descending order as follows. Let $\bar{A}' = \{a \mid a \in A,\ vc_a > 0.85, \text{and road "a" has at least 2 lanes}\}$. Let $\bar{A}$ be the vector of the link indices in $\bar{A}'$ sorted in descending order of M_a which is the merit index of link as defined below (note that the size of $\bar{A}$ is denoted by $|\bar{A}|$)

$$M_a = \frac{vc_a \cdot \frac{\bar{\bar{x}}_a}{x_a} \cdot (x_a + \bar{\bar{x}}_a)}{c_a} \quad \text{for } a \in \bar{A} \tag{10}$$

where $\bar{A}$ is the ordered list of the candidate roads (sorted in descending order) , vc_a is the volume per capacity ratio of the link

. As noted before, it is evident that taking away space from already underutilized roads (i.e., low vc_a) to the transit bears no additional advantages. Hence, we consider a threshold of $vc_a > 0.85$ equivalent to level of service E which is regarded as the "working at capacity condition" (HCM2010, 2010).

This threshold or equivalently the LOS E is the approximate point at which the speed of the traffic flow suddenly drops (see (HCM2010, 2010) Exhibit 11-6 p. 11-8 and Exhibit 11-15, p. 11-20). As noted before, the aim is not to lay out a transit lane on uncongested roads (i.e., $vc_a < 0.85$). Moreover, in order to keep the connectivity level of the network intact, only roads with at least two lanes (per direction) are designated for the transit lane. Having said that, a two-lane road with $vc_a > 0.5$ could easily become over saturated with the implementation of a transit lane.

It is important to reiterate that the mandate of priority planning is to give priority to mass transit like buses even at the cost of more delay for private cars, to encouraging them to shift to the public transport. However, without any modal shift - according to the BP- the example road could be found of no interest by private cars, that is, its traffic volume could come down to zero. In other words, these private cars may have found other shorter paths. This is the beauty of the BP. Of course, there might be some cases in which some (uncongested) roads become highly congested, but, by minimizing the objective function the overall performance of the networks will not deteriorate.

According to condition (10), the more congested a road is, the greater the chance it has to be designated as a dedicated transit lane. We added $\bar{\bar{x}}_a / x_a$ to ensure that the road with a high percentage of transit flow is likely to receive transit priority. Between two roads with the same volume-capacity ratios, the one that carries more traffic is more likely to be designated as a dedicated transit lane. That is why we added the term $x_a + \bar{\bar{x}}_a$ in condition (10). Nonetheless, the merits of the roads are normalized by their respective costs. It is a greedy way to push more cost efficient roads to the top of the merit list (the list in descending order). The numerical result shows that the above index does a good job, such that the projects selected in the final optimum solutions are among the top ones in the sorted list.

Transit lanes are just like auto lanes with pce vehicles running in them. Nevertheless, the BP could still occur when we take the lanes off the existing links as bus lanes. We are not removing the whole link, but just removing some lanes and re-using them in another form. To address such concerns, it is worth noting that the aim of this paper is not to eradicate the Braess paradox, rather the aim is to make the best out of the likely existing BP to "promote" and advocate the public transport ridership. In other words, despite all our effort, we may end up reaching a situation in which a number of roads are designated to cede a lane to the public transport while the BP still exists among them. However, we have already achieved to improve the overall performance of the network compared to the existing situation, even though the BP may still exist.

Do we need to completely block the Braess-tainted links? This deserves to be a subject for further investigation. A complete road closure is a very sensitive action (not from the traffic point of view, more from politically-vested interests) with implications for land use (business, outlets, shops along the respective road ought to suffer and resist) etc.

It is also important to note that, the proposed methodology provides a pro-public transport network (i.e. flagging some bus lanes throughout the road network) without compromising the integrity, connectedness and performance of the network compared to the do-nothing network. Any other good things – like finding the BP over the rest of the network- can be treated as boons which deserve more investigation. We refer the reader interested to know more about BP detection to (Bagloee et al., 2013).

A similar concern may arise with respect to the way we derive the candidate set (condition (10)). A better approach seems to be to detect the BP automatically and remove them from the network (even not to be used as bus lane). To this end, further to what we discussed above, it is important to highlight two points. First, the BP detection is an extremely difficult problem (Roughgarden and Tardos, 2002), second, there is a practical advantage in our proposed methodology regarding condition (10), as follows. We first come up with a set of candidate roads suspected to be Braess-tainted. The BB algorithm is then launched over this set to identify the best subset. This initial candidate set can also be altered, based on other concerns (for example, traffic authorities might be interested to practice a number of what-if scenarios). To this end, it is widely believed that transportation is largely driven by non-transportation vested interests. Furthermore, a complete road closure –as is the case in the BP detection- is highly controversial. That could jeopardize the whole point of promoting public transport.

- **A tight lower bound**

For the mixed integer programming problem of the TPLDP, given the candidate set $\overline{A}$ (or the binary decision variables) the algorithm initiates from the existing (do nothing) scenario $z_j = (0,0,0,0,0)$ represented by the first node of the tree $(j = 0)$. Each node in the tree represents either a partial or complete solution. For example, if there are five binary variables, solution (0,1,0,2,2) represents the situation in which only the first three components are determined with values of 0/1 and the last two, represented by 2, are as yet unspecified; hence, it is a "partial solution". Each time a node z is added to the tree, a lower bound based on the system optimal (SO) follows and the total travel time (objective function (1)) is evaluated (LeBlanc, 1975). Therefore, all the free binary variables "2" are set to "0" and SO flow on the network is computed.

The SO flow for the respective network of z_j is computed and the total travel time corresponding to objective function (1) is set as a lower bound. The UE flow for the network of z_j is also computed and the corresponding total travel time is saved as the upper bound and is called the incumbent value. As the tree expands, the incumbent value takes the objective value (total travel time, the objective function (1)) of UE flow of the best solution found. In other words, the incumbent value is the minimum of the upper bounds.

One of the key factors contributing to the efficacy of the algorithm is rooted in how narrow the distance is between the lower bounds and the incumbent values. It is important to note that the lower bound and the incumbent values are calculated based on SO and UE flows, respectively. The ratio of the travel time of the UE flow to SO flow, called "price of anarchy," can be as high as 2.15 (Roughgarden and Tardos, 2002). In this section, we propose a heuristic procedure to relax the SO flow in order to bridge such a wide gap between SO and UE.

The SO flow can be easily computed using commercial transport planning software by replacing the delay function in the objective function of the UE flow (objective function (4)) to (Sheffi, 1985):

$$\tilde{t}_a(x_a + \bar{\bar{x}}_a) = t_a(x_a + \bar{\bar{x}}_a) + x_a.\frac{\partial t_a(x_a + \bar{\bar{x}}_a)}{\partial x_a}\,, \tag{11}$$

where, if $t_a(x_a + \bar{\bar{x}}_a)$ is considered as the cost of traveling on road $a \in A$ (a BPR delay function), then $\tilde{t}_a(x_a + \bar{\bar{x}}_a)$ is known as the marginal cost of using the respective road. The wide gap between the SO and the UE emerges from the second term on the right-hand side of the equation (11) which is the additional externality cost imposed on the users. The two functions $t, \tilde{t}$ show a similar behavior as long as the volume is below capacity. As the volume reaches (or exceeds) the capacity level, the externality cost increases rapidly, hence it results in a wide gap between the SO and UE flows. Since the capacity of the delay function is not restricted, in theory, the volume and hence the delay as well as the marginal cost can increase to infinity, which results in a much wider gap between SO and UE. In order to decrease such an unrealistic gap, we propose $0 \le \alpha \le 1$ as a coefficient in the externality term:

$$\tilde{t}_a(x_a + \bar{\bar{x}}_a) = t_a(x_a + \bar{\bar{x}}_a) + \alpha \ . \ x_a . \frac{\partial t_a(x_a + \bar{\bar{x}}_a)}{\partial x_a} \tag{12}$$

As α reaches zero the SO moves closer towards the UE and the gap vanishes. It is worth noting that the α addresses the unfortunate tradeoff between computational time and accuracy of the algorithm. The lower the α, the more quickly it leads to a less accurate solution. The value of α can be chosen at the user's discretion depending on the current computational technology and the affordability of the computational time. In other words, α is a valve in the hand of the modeler based on which, the accuracy of the results along with the computational time can be adjusted depending on the size of the network, available computational technology and the strategic value of the final solution. Though, the α, in fact, simplifies the problem to an SO relaxation, the validity of the results especially for real-life networks are strongly upheld. Recent studies have shown that the difference the UE and SO traffic patterns stand in close proximity such that for the case of the city of Chicago the difference does not exceed 6% across the entire network (Boyce and Xiong, 2004; Zheng and Boyce, 2011). In other words, the α reflects on the observation that, for the real life road networks, the SO traffic pattern stands close to the UE traffic pattern. As the result, the α can be considered as an engineering way to tap into this revelation to our advantage enabling us to address the real-sized networks. In the next section, we discuss the values of the α for the case study undertaken.

5. Numerical demonstration

Real size transportation data for Winnipeg, Canada which is widely available in the literature (Bar-Gera, 2016) is used to evaluate the proposed methodology. The case study is comprised of 154 zones, 903 nodes, 2995 directional links and 133 transit lines. Total hourly car and transit passenger demands are 56,219 and 18,211, respectively. As for the computational technology we employ a desktop computer with Intel(R) Core(TM) 3.40 GHz and 16.0 GB RAM. The algorithm is written using Visual Basic linked with MS-Excel as an interface and MS-Access to handle the data efficiently; it is also synchronized to EMME3 to carry out bimodal traffic assignments.

Table 1 shows candidate roads sorted according to their merit indices (condition (10)), adding up to 15.27 km of roads to be considered in the analysis. We consider the length of the roads as the corresponding costs $(c_a, a \in \bar{A})$ and the total budget is $B = 10.00$ (in units of length). As for convergence of private traffic assignment we use a relative gap of 1% which is proven empirically to be close to an acceptable level of equilibrium assignment (INRO, 2009). Note that the traffic assignment problem is solved iteratively and in each iteration the relative gap –as a termination condition- is computed as the difference between the total travel times calculated, based on the currently used paths and current shortest paths, divided by the former. At each iteration, the traffic assignment including private and transit quickly converges within 3 seconds. The total travel time of the private cars and the transit passengers in the do-nothing scenario is calculated as 978634.6 car-minutes and 466665.7 passenger-minutes, respectively.

We started the algorithm with a meagre value for α (i.e. $\alpha = 0.010$). As shown in Table 2 the algorithm terminates within almost two hours and the best solution found entails three links (links 1, 3 and 11 of Table 2) with a total length of 0.24 km out of a total budget of 10.00 km. It is worth noting that the transit priorities are sought over the Braess-tainted roads, and the available budget of 10.00 km roads does not imply that potentially there are 10 km Braess tainted roads available. That is the reason why less than the entire budget has been used. The total travel time (private and public modes) of this solution became 972671.1 car-minutes and 463865 (passenger-minutes), respectively, which is 0.61% improvement compared with those of the do-nothing scenario.

The algorithm was able to shrink the search domain over the branch-and-bound tree on 1,290 occasions in which the lower bound stood above the incumbent values.

Table 1 Candidate links to be considered as transit priority lanes; data for Winnipeg, Canada

no	I_node	J_node	length* (km)	number of lanes	free flow speed	capacity per lanes	hourly car volume	hourly transit volume (in car equivalency)	volume per capacity	sorting Index
1	1046	1045	0.07	4	35	375	1419.52	328.529	1.17	6761.98
2	1047	1050	0.07	4	35	375	1002.49	283.464	0.86	4467.29
3	1050	1047	0.07	4	35	375	1685.39	199.892	1.26	4024.8
4	1044	1043	0.1	4	35	375	1258.42	304.892	1.04	3939.12
5	1047	1046	0.09	4	35	375	1771.47	223.529	1.33	3720.08
6	937	948	0.1	4	35	375	1001.74	282.24	0.86	3111.15
7	1041	1040	0.07	4	35	375	1466.54	167.877	1.09	2913.32
8	931	937	0.1	4	35	375	1030.56	262.24	0.86	2829.15
9	1051	1050	0.11	4	35	375	1685.39	209.892	1.26	2703.63
10	1045	1044	0.16	4	35	375	1258.42	304.892	1.04	2461.95
11	917	931	0.1	4	35	375	1200.65	216.163	0.94	2397.76
12	1042	1025	0.12	4	35	375	1023.83	254.38	0.85	2249.55
13	1043	1042	0.18	4	35	375	1261.47	304.892	1.04	2187.37
14	901	917	0.1	4	35	375	1196.33	194.734	0.93	2105.82
15	1020	1019	0.05	2	25	200	286.331	83.0119	0.92	1970.24
16	1010	1009	0.06	4	35	375	1747.37	91.6667	1.23	1977.75
17	899	898	0.08	4	35	625	2056.45	129.375	0.87	1495.47
18	1008	1007	0.07	4	35	375	1528.51	91.6667	1.08	1499.1
19	606	605	0.14	4	50	625	3019.8	139.375	1.26	1312.27
20	967	980	0.1	2	25	200	250.28	98.5119	0.87	1194.4
21	947	967	0.1	2	25	200	249.761	98.5119	0.87	1195.1
22	411	410	0.1	3	40	875	2843.67	89.3647	1.12	1032.34
23	1034	1035	0.09	3	40	875	2383.62	89.3647	0.94	968.36
24	1011	1010	0.12	4	35	375	1609.37	91.6667	1.13	912.36
25	412	411	0.1	3	40	875	2631.72	77.3647	1.03	820.28
26	1037	1038	0.17	4	35	375	1429.08	114.365	1.03	748.37
27	1009	1008	0.17	4	35	375	1512.47	108.333	1.08	737.53
28	1036	1037	0.18	4	35	375	1418.92	114.365	1.02	700.3
29	1041	1042	0.12	4	35	375	1249.88	84.5119	0.89	669.18
30	605	604	0.24	4	50	625	2426.03	139.375	1.03	632.51
31	414	973	0.12	2	45	1250	2295.62	76.6667	0.95	627.21
32	607	606	0.37	4	50	625	3205	159.174	1.35	609.61
33	410	1034	0.15	3	40	875	2383.62	89.3647	0.94	581.01
34	604	603	0.25	4	50	625	2426.03	129.375	1.02	556
35	1035	1036	0.22	4	35	375	1870.51	89.3647	1.31	557.55
36	170	169	0.23	3	50	625	2478.54	85.8036	1.37	528.78
37	1053	1052	0.24	4	35	375	1180.54	129.089	0.87	519.12
38	608	607	0.22	4	50	625	2993.71	84.799	1.23	487.53
39	1012	1011	0.17	4	35	625	2384.09	76.6667	0.98	456.17
40	415	414	0.23	2	40	875	2169.41	76.6667	1.28	441.75
41	166	165	0.2	2	55	1250	2243.53	85.8036	0.93	414.25
42	603	602	0.37	4	50	625	2655.43	129.375	1.11	407.03
43	600	599	0.36	4	50	625	2331.27	129.375	0.98	371.73
44	599	600	0.36	3	50	625	1677.41	129.375	0.96	371.61
45	887	899	0.33	4	35	625	2056.45	129.375	0.87	362.54
46	167	166	0.23	2	55	1250	2243.53	85.8036	0.93	360.21
47	601	600	0.42	4	50	625	2199.21	129.375	0.93	303.33
48	601	602	0.58	3	50	625	1759.88	129.375	1.01	241.85
49	602	601	0.58	4	50	625	2248.88	129.375	0.95	224.1
50	165	1055	0.37	2	55	1250	2243.53	85.8036	0.93	223.92
51	304	412	0.38	3	40	875	2631.72	77.3647	1.03	215.86
52	973	1012	0.35	2	45	1250	2295.62	76.6667	0.95	215.05
53	437	436	0.46	3	50	625	2003.75	76.6667	1.11	192.08
54	175	174	0.45	3	50	625	1711.39	79.6271	0.96	177.77
55	441	442	0.51	3	50	625	1912.45	78.5417	1.06	169.95
56	442	441	0.51	3	50	625	1822.49	78.5417	1.01	162.25
57	1055	1059	0.75	2	55	1250	2243.53	85.8036	0.93	110.47
58	436	423	0.82	3	40	875	2311.83	76.6667	0.91	87.9
59	423	415	0.89	3	40	875	2495.26	76.6667	0.98	87.01
60	423	436	0.82	3	50	625	1526.17	76.6667	0.85	83.46

*Total length is 15.27 km.

In the next runs, the algorithm was tested for two additional values of α (i.e. $\alpha = 0.015$ and $\alpha = 0.020$) as shown in Table 2. As can be seen, the computational times for each value of the α have been reported based on which the one pertaining to the α of $\alpha = 0.020$ lasted almost three days. In these two runs of the α s , the same solution was obtained, in which the total travel time (private and public modes) became 971429.3 car-minutes, and 463256.1 (passenger-minutes), respectively, equivalent to approximately 0.74% improvement compared with those of the do-nothing scenario.

In terms of the roads identified for transit priority lanes (roads 1 to 15 and 29 of Table 2), one can easily observe that they are chosen consistently from the top of the sorted list in Table 1. It indicates that the notion of sorting the candidate roads on the merit basis proposed in condition (10) is obviously working. Furthermore, the depleted budget is 1.61 km which accounts for 16% of the available budget. It is evident that an increase of 1% in the value of α resulted in a drastic increase in the computation time. This implies that the value of α influences the size of the solution domains strongly. Furthermore, Figure 1 shows the computational cost as well as the incremental improvements of the value of the objective function over a range of the values of α. That is, higher values of α result in better solutions (i.e., lower values of the objective functions), but at the expense of a higher computational time.

Figure 2 demonstrates graphically the roads identified to provide one lane dedicated to transit. Apart from a few sporadic roads, the transit priority lanes are topographically consistent.

Table 2 Numerical result for data for Winnipeg, Canada

α	No of times lower bound higher than incumbent values	Computation time (hours)	The string of solution found (1: to-build, 0: no-build)	Total Travel Time (car-minutes)	Transit travel time (passenger - min)	Budget used out of 10
0.010	1290	1.874	101000000010000000000000000000000000000000000000000000000000	972671.1	463865.7	0.24
0.015	4995	5.297	111111111111111000000000000010000000000000000000000000000000	971429.3	463256.1	1.61
0.020	31503	68.719	111111111111111000000000000010000000000000000000000000000000	971429.3	463256.1	1.61

It is worth noting that, given the current level of congestion (i.e. the do-nothing scenario represented by the total travel time of 978634.6), the addition of the transit priority lanes not only did not degrade the traffic circulation, but in fact, it improved –though slightly - the overall traffic circulation (0.74% reduction in the total travel time). It shows clearly that the initial idea of the BP exists and it worked in our favor as expected.

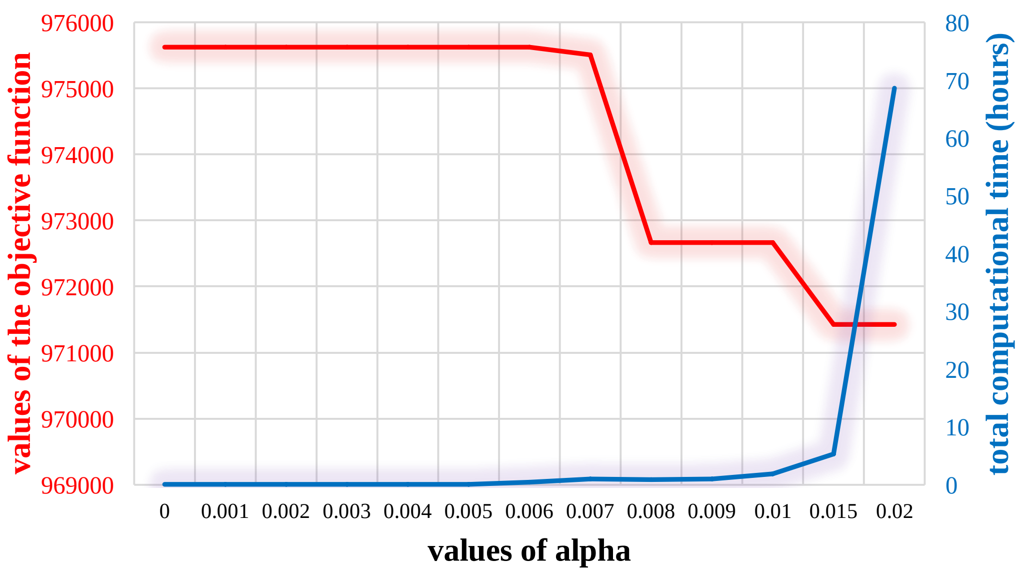


Figure 1 Impact of the α on the computational time and the objective function

Although it sounds minuscule, the 0.74% improvement is, in fact, a boon. We endeavored to show that bus lanes can be added to the network at no adverse cost. However, the consensus in the literature is to add bus lanes even at the cost of degrading the network for the private modes –sometimes intentionally- to encourage them to shift to the public transport. Nevertheless, the 0.74% hourly saving time -if factored to the day and year and total magnitude of travel demand- would become a significant improvement. Moreover, in examples of road network design (or road constructions), the amount of expected improvement in the saving time is around only a few percent. For instance (Poorzahedy and Rouhani, 2007) reported only 2% improvement as the result of a massive investment on undertaking 10 mega grade-separation projects (see Table 6 of (Poorzahedy and Rouhani, 2007)) in a real life example, whereas the improvement of 0.74% which is equivalent to one-third of 2%, comes at zero cost.

With respect to the BP, the algorithm sought mis-utilized capacity (even in the congested area), to be taken away from the private mode and to be used exclusively by the transit mode. Although the idea of providing priority to the transit is appealing, there is certainly a level above which the overall performance of the network (private and public) will deteriorate. The performance of the network was measured as the total travel time/cost formulated in the objective function (1) and was referred to as the incumbent value. Since the algorithm started with the incumbent value of the do-nothing scenario, in the end, the algorithm did not render any solution worse than that of the do-nothing scenario. For the greater cause of transit priority, should a slight deterioration of the private mode be acceptable, one can re-launch the algorithm with a slightly higher incumbent value. Therefore, more links ought to be found as transit priority lanes. Such measures can be strongly justified in light of possible shifts in travel demand from private to transit if greater priority and incentive are provided for public transport. This brings us to set out new areas for further studies which are discussed in the next section.

6. Conclusions

In this study, we aimed to enhance the attractiveness of the transit system by providing transit priority lanes at no cost, nor additional burden to the private mode. To this end, we were motivated by the Braess Paradox in seeking roads for which closure will counterintuitively improve overall traffic circulation. Accordingly, instead of complete closure, we took away a lane for the transit system and left the rest of the space for the private mode in order to maintain the same level of network connectivity. This problem was formulated as a bilevel, nonlinear programming problem mixed with binary decision variables which are proven to be extremely intractable for large-sized networks. At the upper level, the total system cost (or total travel time) is minimized, while bimodal traffic assignment is taken into account at the lower level.

In order to address the scalability of the methodology, we proposed a greedy and RAM-efficient branch and bound algorithm tailored to large-sized networks. The algorithm was coded using MS Office applications (Access, Excel) synchronized with a commercial transport planning software (EMME 3) with no need for any additional optimization solvers targeting the need of the industry and the practitioners.

In the first phase, a set of roads deemed appropriate for candidacy as transit priority lanes were identified. To this end, criteria such as current transit ridership, congestion level, and even costs pertaining to implementation of a transit lane were considered (condition (10)). In the second phase, a subset of the candidate set was sought using the proposed BB algorithm. In doing so, based on traffic characteristics, a merit index for the candidate projects was calculated. The roads were then sorted in descending order from the most likely to be the best selection for designation as transit priority lanes. The tree structure of the BB is built on the sorted list of candidates. The branching is also done based on the sorted list from the top, descending steeply into the last possible candidate, subject to budget constraints or weak lower bounds. This would help with the use of extremely small RAM space which is a decisive factor in handling large-sized networks. Such a simple rule makes the search over the tree quite smooth, with no effort required to remember the structure of the rapidly growing tree. Subsequently, this offers an ideal leverage for dealing with large-sized networks.

As advised by LeBlanc (1975), the lower bound value was calculated based on system optimal (SO) traffic flow (see also Appendix B). In cases in which no lower bound is found to be higher than the incumbent value, the wide gap between SO and UE flow may affect the numerical affordability of the BB adversely. We, therefore, proposed inserting the α coefficient in the marginal delay function to bridge the gap. The value of α controls the tradeoff between accuracy of the solution and computational time easily. Given the fact that we tie up the proposed algorithm to the Braess Paradox, it may validly be asked how the method would perform if it is applied to a network which is already designed to prevent BP from happening? To answer this question let us first underscore the primary aim of the transit priority, that is to give priority to the public transport-even at the cost of leaving less space to private cars- to encourage people to shift to public transport. Laying out the bus lanes on the shoulder of the Braess Paradox is a boon to the primary cause which is a conservative approach to bus priority. Nevertheless, in the previous studies the bus lanes are added irrespective of maintaining the same level of service for the private mode. In the case of dealing with a BP-free network the proposed methodology can still be applied by launching the branch-and-bound with a higher incumbent value (say infinity; as noted before the initial incumbent value is the total travel time of the do-nothing or existing network). As a result of designating some lanes for the public transport, private cars are inevitably faced with longer travel times.

We evaluated the algorithm using real data for Winnipeg, Canada. The best solution comprises 1.61 km of transit priority lanes primarily located in the central business district. It is important to note that change of the transit network structure would change the bus frequencies and network, but we are ignoring them here. Moreover, we should note that any change in the transit network would change the signal phases and timing. In addition, giving priority to the transit is supposed to entice more ridership (modal shift from private modes and is aimed at doing so). As the result, a worthy line of research is to develop a combined model in which modal split and traffic assignment are synchronized to take the mutual changes (network vs demand) into account fully. Accordingly, the proposed algorithm can be further improved on several fronts as follows: (i) in this study the travel demand matrices for both private and transit modes were assumed to be fixed. Given the intention of transit priority to make the transit mode more appealing, the proposed methodology can be further improved to consider flexible travel demand and hence possible shift from private mode to transit mode. (ii) The concept of transit priority lanes would work more efficiently in synchronization with transit signal priorities. To this end, road delays functions to adjust the priority signal settings become non-separable functions. This gives rise to path-based traffic assignment methods such as complementarity and variational inequality methods (Aashtiani, 1979; Nagurney, 1998) which is still an evolving subject in the literature. (iii) The possible spare capacity of the transit lanes can provide an opportunity for promoting car-sharing schemes or high occupancy vehicles (HOV) as well as the bicycle mode. The model can be further improved by considering the variation of travel demand over time, as well as in response to changes in the network.

Figure 2 Winnipeg transport network and selected transit priority lanes

Phased-implementation of a transit priority scheme to incur the least temporal disruption to traffic circulation is also of highest importance. To this end one my need to go extra mile to prioritize the implementations of the transit priority lanes which calls for an interesting problem of prioritization/scheduling (Bagloee and Asadi, 2015; Bagloee and Tavana, 2012). In light of real-time data, big data, sensor revolution, and the internet of things, some scholars advocate moving toward the dynamic traffic assignment (DTA) that is to include time variation features of the traffic. The DTA is based on the fundamental diagram – a method derived from traffic flow theory- to model the congestion with a greater realism and fidelity. Furthermore, one of the main drawbacks of the (static) traffic assignment method, regardless of having a priority lane, is a lack of considering vehicle-to-vehicle interaction (for example in a one-by-one road, if a car stops, all other should stop, too). To this end, alternative methods are the DTA –or a more disaggregated model- microsimulation which is based on the car the following method. Though the reward is enormous, the task is highly challenging, attributable to some theoretical hardship as well as computational costs. Nevertheless, the DTA seems to be the future, as the result, integration of the DTA in the proposed methodology is a deserving thread of research. Safety issues arising from transit priority lanes in the CBD (Bagloee and Asadi, 2016), deserve to be further investigated in future studies.

Acknowledgment

The authors are indebted to Prof Voss the editor-in-chief and four anonymous reviewers for their meticulous and comprehensive comments. The authors would like to thank Prof Peter Thomson of the Monash University, Australia, for his useful comments.

References:

Aashtiani, H.Z. (1979) The multi-modal traffic assignment problem. PhD dissertation, Massachusetts Institute of Technology.

Aashtiani, H.Z., Poorzahedy, H. (2004) Braess' phenomenon in the management of networks and dissociation of equilibrium concepts. *Transportation planning and technology* 27, 469-482.

Achterberg, T., Wunderling, R. (2013) Mixed integer programming: Analyzing 12 years of progress. *Facets of Combinatorial Optimization*. Springer, pp. 449-481.

Bagherian, M., Mesbah, M., Ferreira, L. (2015) Using delay functions to evaluate transit priority at signals. *Public Transport* 7, 61-75.

Bagloee, S.A., Asadi, M. (2015) Prioritizing road extension projects with interdependent benefits under time constraint. *Transportation Research Part A: Policy and Practice* 75, 196-216.

Bagloee, S.A., Asadi, M. (2016) Crash analysis at intersections in the CBD: A survival analysis model. *Transportation Research Part A: Policy and Practice* 94, 558-572.

Bagloee, S.A., Ceder, A., Tavana, M., Bozic, C. (2013) A heuristic methodology to tackle the Braess Paradox detecting problem tailored for real road networks. *Transportmetrica A: Transport Science* 10, 437-456.

Bagloee, S.A., Sarvi, M., Patriksson, M. (2016a) A Hybrid Branch-and-Bound and Benders Decomposition Algorithm for the Network Design Problem. *Computer-Aided Civil and Infrastructure Engineering*, http://dx.doi.org/10.1111/mice.12224.

Bagloee, S.A., Sarvi, M., Wallace, M. (2016b) Bicycle lane priority: Promoting bicycle as a green mode even in congested urban area. *Transportation Research Part A: Policy and Practice* 87, 102-121.

Bagloee, S.A., Tavana, M. (2012) An efficient hybrid heuristic method for prioritising large transportation projects with interdependent activities. *International Journal of Logistics Systems and Management* 11, 114-142.

Bar-Gera, H. (2016) Transportation Network Test Problems. http://www.bgu.ac.il/~bargera/tntp/; Accessed by Dec. 1, 2016.

Basso, L.J., Guevara, C.A., Gschwender, A., Fuster, M. (2011) Congestion pricing, transit subsidies and dedicated bus lanes: Efficient and practical solutions to congestion. *Transport Policy* 18, 676-684.

Beckmann, M., McGuire, C., Winsten, C.B. (1956) *Studies in the Economics of Transportation*, Yale University Press, New Haven, CT.

Ben-Ayed, O., Blair, C.E. (1990) Computational Difficulties of Bilevel Linear Programming. *Operations Research* 38, 556-560.
Bixby, R.E. (2012) A brief history of linear and mixed-integer programming computation. *Optimization Stories* ed Grötschel, M. Deutsche Mathematiker-Vereinigung, Bielefeld, pp. 107–121.
Boyce, D., Xiong, Q. (2004) User-optimal and system-optimal route choices for a large road network. *Review of network Economics* 3, 371-380.
Ceder, A. (2015) *Public Transit Planning and Operation: Modeling, Practice and Behavior*. CRC Press.
Chen, Q. (2015) An optimization model for the selection of bus-only lanes in a city. *PloS one* 10, e0133951.
Cova, T.J., Johnson, J.P. (2003) A network flow model for lane-based evacuation routing. *Transportation research part A: Policy and Practice* 37, 579-604.
D'Ambrosio, C., Lodi, A. (2013) Mixed integer nonlinear programming tools: an updated practical overview. *Annals of Operations Research* 204, 301-320.
De Cea, J., Fernández, J.E., Dekock, V., Soto, A. (2005) Solving network equilibrium problems on multimodal urban transportation networks with multiple user classes. *Transport Reviews* 25, 293-317.
Diab, E.I., El-Geneidy, A.M. (2013) Variation in bus transit service: understanding the impacts of various improvement strategies on transit service reliability. *Public Transport* 4, 209-231.
Eichler, M., Daganzo, C.F. (2006) Bus lanes with intermittent priority: Strategy formulae and an evaluation. *Transportation Research Part B: Methodological* 40, 731-744.
Fang, Y., Chu, F., Mammar, S., Che, A. (2013) An optimal algorithm for automated truck freight transportation via lane reservation strategy. *Transportation Research Part C: Emerging Technologies* 26, 170-183.
Fang, Y., Chu, F., Mammar, S., Che, A. (2014) A cut-and-solve-based algorithm for optimal lane reservation with dynamic link travel times. *International Journal of Production Research* 52, 1003-1015.
Fisher, M.L. (2004) The Lagrangian relaxation method for solving integer programming problems. *Management science* 50, 1861-1871.
Florian, M., Morosan, C.D. (2014) On uniqueness and proportionality in multi-class equilibrium assignment. *Transportation Research Part B: Methodological* 70, 173-185.
Fontaine, P., Minner, S. (2014) Benders decomposition for discrete–continuous linear bilevel problems with application to traffic network design. *Transportation Research Part B: Methodological* 70, 163-172.
Gao, Z., Wu, J., Sun, H. (2005) Solution algorithm for the bi-level discrete network design problem. *Transportation Research Part B: Methodological* 39, 479-495.
Geroliminis, N., Zheng, N., Ampountolas, K. (2014) A three-dimensional macroscopic fundamental diagram for mixed bi-modal urban networks. *Transportation Research Part C: Emerging Technologies* 42, 168-181.
Guler, S.I., Cassidy, M.J. (2012) Strategies for sharing bottleneck capacity among buses and cars. *Transportation research part B: methodological* 46, 1334-1345.
Guler, S.I., Gayah, V.V., Menendez, M. (2016) Bus priority at signalized intersections with single-lane approaches: A novel pre-signal strategy. *Transportation Research Part C: Emerging Technologies* 63, 51-70.
Guler, S.I., Menendez, M. (2015) Pre-signals for bus priority: basic guidelines for implementation. *Public Transport* 7, 339-354.
Hadas, Y., Ceder, A. (2014) Optimal Connected Urban Bus Network of Priority Lanes. *Transportation Research Record: Journal of the Transportation Research Board*, 49-57.
HCM2010 (2010) Highway Capacity Manual 2010. *National Academy of Sciences. Yhdysvallat*.
INRO (2009) EMME3 v 3.2. *EMME3 User's Guide* 3.2 ed, Montreal, Quebec, Canada.
Khoo, H.L., Teoh, L.E., Meng, Q. (2014) A bi-objective optimization approach for exclusive bus lane selection and scheduling design. *Engineering Optimization* 46, 987-1007.

LeBlanc, L.J. (1975) An Algorithm for the Discrete Network Design Problem. *Transportation Science* 9, 183-199.

Li, S., Ju, Y. (2009) Evaluation of bus-exclusive lanes. *Intelligent Transportation Systems, IEEE Transactions on* 10, 236-245.

Liu, R., Van Vliet, D., Watling, D. (2006) Microsimulation models incorporating both demand and supply dynamics. *Transportation Research Part A: Policy and Practice* 40, 125-150.

Liu, Z., Meng, Q. (2012) Bus-based park-and-ride system: a stochastic model on multimodal network with congestion pricing schemes. *International Journal of Systems Science* 45, 994-1006.

Mesbah, M., Sarvi, M., Currie, G. (2011a) Optimization of Transit Priority in the Transportation Network Using a Genetic Algorithm. *Intelligent Transportation Systems, IEEE Transactions on Intelligent Transportation Systems* 12, 908-919.

Mesbah, M., Sarvi, M., Ouveysi, I., Currie, G. (2011b) Optimization of transit priority in the transportation network using a decomposition methodology. *Transportation Research Part C: Emerging Technologies* 19, 363-373.

Mirchandani, P.B., Li, J.-Q., Hickman, M. (2010) A macroscopic model for integrating bus signal priority with vehicle rescheduling. *Public Transport* 2, 159-172.

Nagurney, A. (1998) *Network economics: A variational inequality approach*. Springer.

Nagurney, A. (2010) The negation of the Braess paradox as demand increases: The wisdom of crowds in transportation networks. *EPL (Europhysics Letters)* 91, 48002.

Poorzahedy, H., Rouhani, O.M. (2007) Hybrid meta-heuristic algorithms for solving network design problem. *European Journal of Operational Research* 182, 578-596.

Roughgarden, T., Tardos, É. (2002) How bad is selfish routing? *Journal of the ACM (JACM)* 49, 236-259.

Sakamoto, K., Abhayantha, C., Kubota, H. (2007) Effectiveness of bus priority lane as countermeasure for congestion. *Transportation Research Record: Journal of the Transportation Research Board*, 103-111.

Sheffi, Y. (1985) *Urban transportation networks: equilibrium analysis with mathematical programming methods*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey

Smith, N., Hensher, D. (1998) The future of exclusive busways: the Brazilian experience. *Transport Reviews* 18, 131-152.

Spiess, H. (1984) Contributions à la théorie et aux outils de planification des réseaux de transport urbain. Montréal: Université de Montréal, Centre de recherche sur les transports.

Spiess, H. (1990) Technical note—Conical volume-delay functions. *Transportation Science* 24, 153-158.

Spiess, H., Florian, M. (1989) Optimal strategies: a new assignment model for transit networks. *Transportation Research Part B: Methodological* 23, 83-102.

Sun, X., Lu, H., Fan, Y. (2014) Optimal Bus Lane Infrastructure Design. *Transportation Research Record: Journal of the Transportation Research Board*, 1-11.

Tse, L.Y., Hung, W.T., Sumalee, A. (2014) Bus lane safety implications: a case study in Hong Kong. *Transportmetrica A: Transport Science* 10, 140-159.

Viegas, J., Lu, B. (2004) The intermittent bus lane signals setting within an area. *Transportation Research Part C: Emerging Technologies* 12, 453-469.

Wang, J., Liu, H., Xie, C. (2016) Transit Network Design with Exclusive Bus Lanes. *Proceedings of Transportation Research Board 95th Annual Meeting*.

Wang, S., Meng, Q., Yang, H. (2013) Global optimization methods for the discrete network design problem. *Transportation Research Part B: Methodological* 50, 42-60.

Wu, P., Che, A., Chu, F. (2013) A quantum evolutionary algorithm for lane reservation problem. *Proceedings of Networking, Sensing and Control (ICNSC), 2013 10th IEEE International Conference on*, pp. 264-268.

Wu, P., Chu, F., Che, A. (2015) Mixed-integer Programming for a New Bus-lane Reservation Problem. *Proceedings of Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference on*, pp. 2782-2787.

Wu, P., Chu, F., Che, A., Shi, Q. (2014) A bus lane reservation problem in urban bus transit network. *Proceedings of Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*, pp. 2864-2869.

Xie, X., Chiabaut, N., Leclercq, L. (2012) Improving Bus Transit in Cities with Appropriate Dynamic Lane Allocating Strategies. *Procedia-Social and Behavioral Sciences* 48, 1472-1481.

Yang, H., Huang, H.-J. (2005) *Mathematical and economic theory of road pricing*.

Yao, J., Shi, F., An, S., Wang, J. (2015) Evaluation of exclusive bus lanes in a bi-modal degradable road network. *Transportation Research Part C: Emerging Technologies* 60, 36-51.

Yao, J., Shi, F., Zhou, Z., Qin, J. (2012) Combinatorial Optimization of Exclusive Bus Lanes and Bus Frequencies in Multi-Modal Transportation Network. *Journal of Transportation Engineering* 138, 1422-1429.

Yingfeng, W., NaiQi, W. (2010) An approximate algorithm for the Lane Reservation Problem in Time Constrained Transportation. *Proceedings of Advanced Computer Control (ICACC), 2010 2nd International Conference on*, pp. 192-196.

YunFei, F., Feng, C., Mammar, S., Che, A. (2011) Iterative algorithm for lane reservation problem on transportation network. *Proceedings of Networking, Sensing and Control (ICNSC), 2011 IEEE International Conference on*, pp. 305-310.

Zhang, L., Yang, H., Wu, D., Wang, D. (2014) Solving a discrete multimodal transportation network design problem. *Transportation Research Part C: Emerging Technologies* 49, 73-86.

Zheng, J., Boyce, D. (2011) Comparison of User-Equilibrium and System-Optimal Route Flow Solutions under Increasing Traffic Congestion, 11-0581. *Proceedings of Transportation Research Board 90th Annual Meeting*.

Zheng, N., Geroliminis, N. (2013) On the distribution of urban road space for multimodal congested networks. *Transportation Research Part B: Methodological* 57, 326-341.

Zhou, Z., Che, A., Chu, F., Chu, C. (2014) Model and Method for Multiobjective Time-Dependent Hazardous Material Transportation. *Mathematical Problems in Engineering* 2014.

Zhou, Z., Chu, F., Che, A., Mammar, S. (2012) A multi-objective model for the hazardous materials transportation problem based on lane reservation. *Proceedings of Networking, Sensing and Control (ICNSC), 2012 9th IEEE International Conference on*, pp. 328-333.

Appendix A: the branch-and-bound

In the following exposition, we adopted the same terminology used by (LeBlanc, 1975). The strings of solutions are either (i) partial, like (0,1,0,2,2) representing the situation in which only the first three components are determined with values of 0/1 and the last two, represented by 2, are as yet unspecified or (ii) complete, representing the situation in which all projects are decided and assigned values 0 or 1 (the previously mentioned partial solution can eventually become any of the following complete solutions (0,1,0,0,0), (0,1,0,0,1), (0,1,0,1,1), (0,1,0,1,1).

- **Lower bounds**

The algorithm initiates by calculating SO flow for the do-nothing scenario (in case of five projects it becomes (0,0,0,0,0)) which renders the lowest possible lower bound value. In the course of branching, whenever two nodes are added to the tree, it is only necessary to compute one lower bound since the other one has already been computed. For example for partial solution $z_j = (0,1,0,2,2)$, at node j of the tree, the branching at the fourth project will end up at two new solutions $z_l = (0,1,0,0,2)$ on the left side and $z_r = (0,1,0,1,2)$ on the right side. The lower bound of z_l which is SO flow of $(0,1,0,0,0)$ has already been calculated when node j was added. Hence it is only necessary to calculate the lower bound of z_r which is the SO flow of $(0,1,0,1,0)$. This process is shown graphically in Figure 3(a).

- **Node selection and branching rules**

At the selected node, a free variable (or undecided project, represented by "2") has to be selected to be assigned 0 and 1, which is called branching. There are some methods to make a move on the tree that sometimes requires solving an additional problem or retrieving the entire database to find what is –hopefully- the best node and branching. Such methods become computationally intensive, as the size of the network increases, Alternatively, based on the merit index we propose the following rules: (i) for node selection; take the deepest node on the branch which has emerged from adding a project successively. Note that at each node, two branches come out: one corresponding to deciding to add a candidate ($y_a = 1$) and the other one not to add ($y_a = 0$). (ii) for branching; select the very next undecided project in the row of the respective partial solutions. As described before (see condition (10)), the projects have already been sorted based on their merits. Hence, it makes sense to go deep into the tree and select the next best project for branching, hoping that the optimum solution lies there. These two simple and intuitive rules obviate any burden of retrieving the information relating to the entire tree. At each node, the algorithm just needs to move forward as much as possible through the $y_a = 1$ branch. In case there is no space to move forward, the algorithm moves only one node back to the previous node and then moves through the $y_a = 0$. Once it reaches a new node, the algorithm proceeds normally, meaning that it first moves through $y_a = 1$ if possible, otherwise it moves through $y_a = 0$. The process goes on until a termination criterion is met. Figure 3(b) shows the gradual buildup of the tree based on the above-mentioned rules for a case consisting of three candidates.

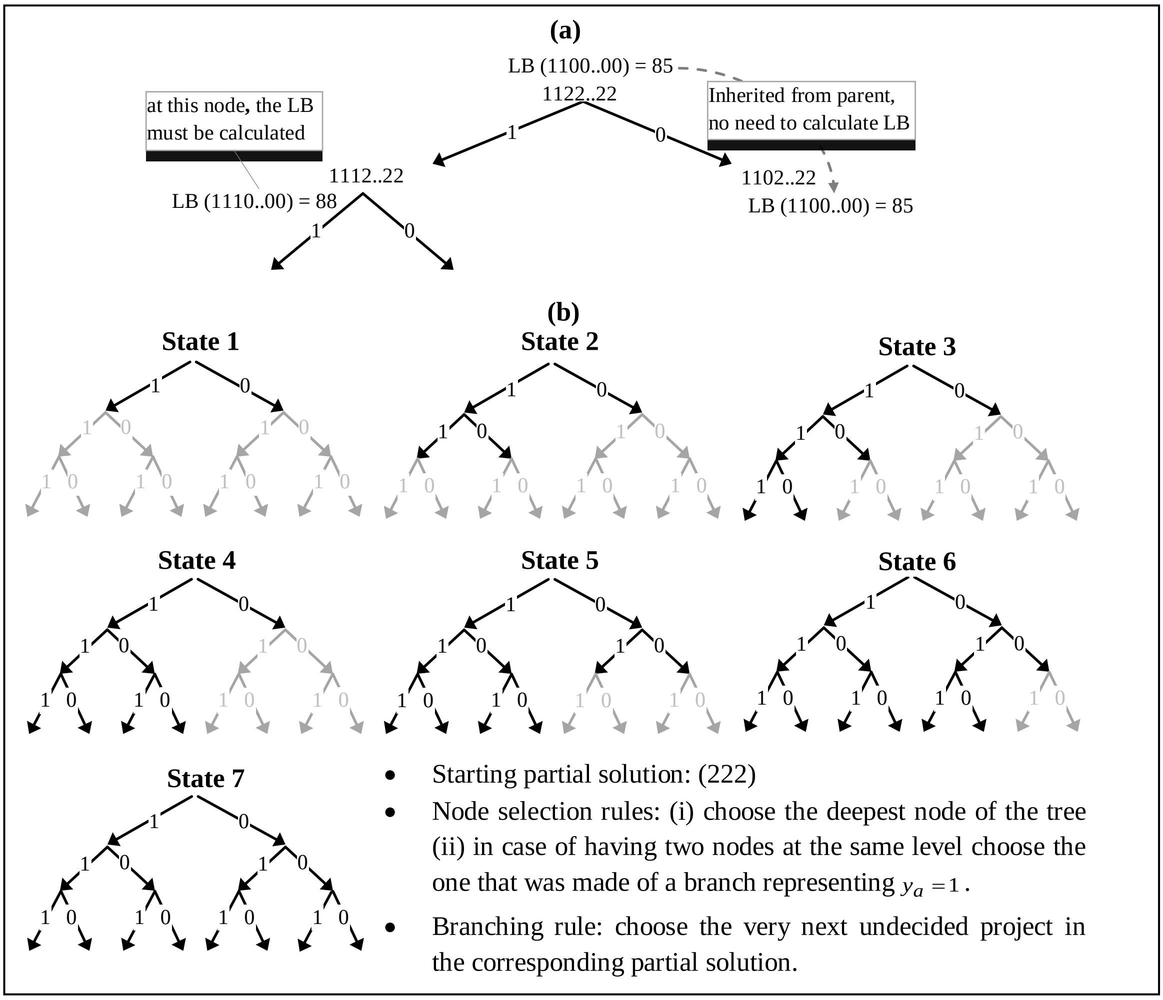


Figure 3 How does the tree structure of the branch and bound algorithm work? (a) proposed node selection and branching rules; (b) efficacy of calculating the lower bound (Bagloee et al., 2016a; Bagloee et al., 2016b)

As the tree structure expands the algorithm does not need to remember the paths already taken, nor the paths ahead. As shown in Figure 3(b), it just needs to know the lower bounds of the nodes on the current path, plus the best solution found so far, which is a string of binary values (0/1), and the corresponding incumbent value. For example, if the current node is (11002), the next move is to process node (11001) followed by the node represented by (11000). For the third move, the algorithm moves three nodes back to reach node (10222). The navigation on the BB's tree is, therefore, memoryless, which is highly advantageous in dealing with networks of large size. In such cases, the rapid expansion of BB tree to extend memorizing the topography and structure of the tree comes at a heavy RAM cost.

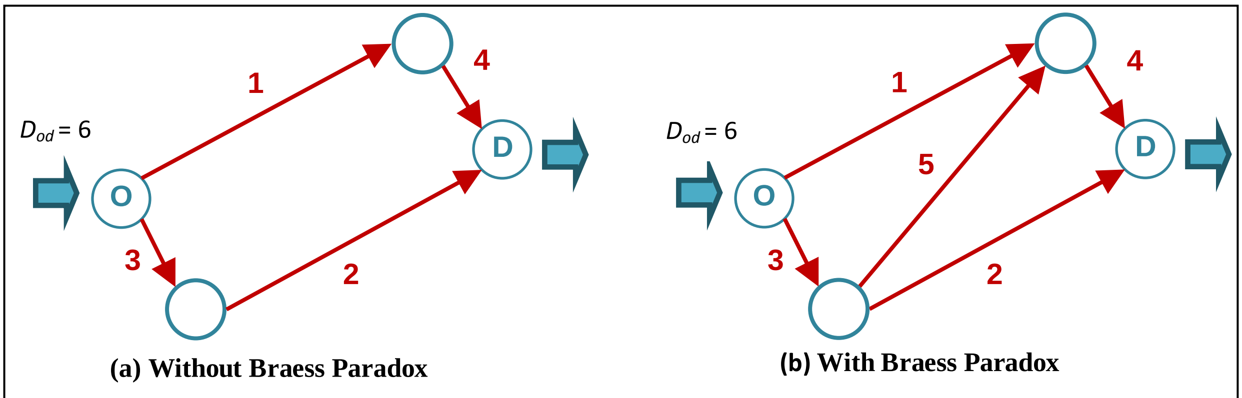
Once the algorithm reaches a new node in the tree, it then determines whether the node stands for partial solutions (some variables are "2": free to be 0 or 1) or complete solutions (all binary variables are determined to be 0 or 1). For complete solutions, the UE flows and the total travel time (upper bound) for the corresponding networks are computed. This solution is then compared with the incumbent solution (the best solution found). The current solution is labeled as the incumbent solution if its upper bound is lower than the incumbent value. For the partial solutions; the SO flow is solved to compute the lower bound on the successors of z_l . Note that the lower bound for z_r which is the same as the predecessor node (z_j) has already been computed.

- **Termination condition**

On three occasions the extension of the tree at a particular node stops or freezes or is called fathoming: (1) reaching the bottom of the tree where there is no free variable, (2) no more project can be added to the branch represented by $y_a = 1$ due to depleted budget, (3) the lower bound is found greater than the incumbent value, then the respective node is deleted from the list of partial solutions and the process continues until no partial solution is left (termination condition).

Appendix B: Braess Paradox example

Our conjecture is as follows: "1" means the respective road has to give away a lane as a bus lane and "0" means not, and the current mixed use lane is not appropriate because of BP, and some 1's (actions) may improve it". Similar to the famous Braess example shown in Figure 4, network (b) with additional road "5" is Braess-tainted compared with the network "a". It is proven that even in the network "b" (where the BP exists) if we charge drivers the marginal costs of the roads, the Braess-tainted road "5" becomes unattractive as if it was never open to the traffic like Network "a". Charging drivers the marginal costs results in SO traffic flow which are equivalent to the minimum possible travel time. Therefore keeping all the roads, even the Braess-tainted roads (like network "B") open to the traffic (that is $(y_a) = 0$), and charging drivers with the marginal cost (SO flow) would result in the minimum possible travel time, which is mathematically a valid lower bound.



Delay Functions: $t_1(x) = t_2(x) = x + 50$; $t_3(x) = t_4(x) = 10x$; $t_5(x) = x + 10$

Figure 4. Braess Paradox where removing link 5 yields a better traffic flow