Optimizing Dedicated Bus Lane allocation in bi-modal networks with dynamic congestion

Dimitrios Tsitsokas

Anastasios Kouvelas

Nikolas Geroliminis

École Polytechnique Fédérale de Lausanne

May 2019

STRC | 19th Swiss Transport Research Conference
Monte Verità / Ascona, May 15 – 17, 2019

École Polytechnique Fédérale de Lausanne

Optimizing Dedicated Bus Lane allocation in bi-modal networks with dynamic congestion

Dimitrios Tsitsokas, Nikolas Geroliminis
Urban Transport Systems Laboratory
École Polytechnique Fédérale de Lausanne
GC C2 390, Station 18, 1015 Lausanne, Switzerland
phone: +41-21-69-32481
fax: +41-21-69-32431
{dimitrios.tsitsokas, nikolas.geroliminis}@epfl.ch

Anastasios Kouvelas
Institute for Transport Planning ans Systems
ETH Zurich
HIL F 37.2, Stefano-Franscini-Platz 5, 8093 Zurich, Switzerland
phone: +41-44-633 66 95
fax: +41-44-633 66 95
anastasios.kouvelas@ivt.baug.ethz.ch

May 2019

Abstract

Increasing urbanization and car ownership rates in developed countries result in hyper-congested cities with long delays in everyday commutes. Since scarcity of space forbids road space expansion, increasing the effectiveness and attractiveness of mass transit seems as a viable way to increase the systems supply. Dedicated Bus Lanes (DBL) provide exclusive space for public transit vehicles and therefore reduce mean travel time as also the congestion-related travel time variance; however, they decrease road capacity for regular traffic that may lead to severe congestion propagation due to the spillback effect. In this work, Variable Neighborhood Search is applied to seek for the optimal DBL plan in terms of total passenger hours travelled. A queueing theory based traffic flow model with proper treatment of spill-backs and traffic signal settings called "Store-and-Forward" is extended to simulate the evolution of traffic congestion in the network. The evaluation of the resulting optimal solution in comparison to state-of-practice solutions through microsimulation can reveal the degree of improvement in terms of decreasing total passenger delay.

Keywords

Dedicated Bus Lanes; Store-and-Forward; Dynamic Traffic Modelling; Variable Neighborhood Search (VNS); Optimization; Bus Priority

1 Introduction

Rapid population growth in many modern cities around the world has caused extensive degradation of mobility in central districts, as a result of high vehicle accumulation in limited road space. In such a situation, maximizing the level of service of mass transit systems seem as a direct way of improving traffic network performance by reducing the number of vehicles in congested areas. Improved mass transit can motivate commuters to use public transit instead of private cars. Many strategies have been developed to deal with this problem, ranging from changes to the infrastructure (dedicated bus lanes, bus rapid transit, e.g. Dion *et al.* (2004), Yao *et al.* (2012),Deng and Nelson (2011)) to applications of smart technologies (traffic signal priority, pre-signals, bus lanes with intermittent priority, as in Eichler and Daganzo (2006) ,Viegas and Lu (2001), Guler *et al.* (2016), Guler and Cassidy (2012). Dedicated Bus Lanes (DBLs) in particular, have seen wide implementation as a way of decreasing travel time for buses by reserving exclusive space for them to move through congestion without long delay. However, many of these measures have non-negligible effects on other vehicles in the system, especially when their demand is high (see Dion *et al.* (2004)) that makes the careful planning of DBL necessary.

Although there has been significant research on the problem of optimizing the allocation schemes of DBLs in traffic networks, in most works found in the literature, the decision making is based on static traffic simulation models that fail to capture the dynamic characteristics of congestion propagation and spill-back effects. As a result, the effectiveness of the proposed solution is restricted to networks experiencing low congestion levels. Dynamic treatment of traffic and integration of operational characteristics even at the planning stage is important for congested urban networks with limited available capacity. This is why addressing the DBL optimal allocation problem requires a framework that can imitate the dynamics of congestion while being easily integrated in an optimization scheme.

The setting of dedicated bus lanes leads to a significant modification of the network performance characteristics, as storage capacity and saturation flow in links with a DBL are decreased. This fact alters the established traffic user equilibrium and may lead to different route patterns and mode choice trends. The interrelation between these effects is also made visible in several studies such as those of Mesbah *et al.* (2011), Bingfeng *et al.* (2017) and Yu *et al.* (2015), where the DBL optimization problem is structured as a bi-level programming model, similar to a leader-follower Stackelberg game; the DBL allocation strategy is the leader problem (system optimum) while mode choice and traffic assignment form distinct optimization subproblems of the follower problem. However, static models are used in these works which fail to imitate the dynamics of congestion and the spillback effect. At the same time dynamic models of route choice create strong computational burden that make the feasibility of an efficient optimization framework very challenging.

In Zheng and Geroliminis (2013), the authors address the DBL setting problem in a multi-modal network in the region level by modeling the congestion dynamics of the complex system via the Macroscopic Fundamental Diagram (MFD) concept, which was validated with experimental results by Geroliminis and Daganzo (2008). In this work, a dynamic traffic model is designed and integrated in an optimization framework with the objective of identifying the optimal fraction of road space that should be allocated to DBLs in every region with the aim of minimizing the total travel delay for passengers. MFD approach was also applied in the work of Gonzales and Daganzo (2012), who examined system optimum and user optimum solutions in a bi-modal network for the morning commute problem. The proposed model shows the decrease in the network storage capacity for cars when DBLs are introduced. Nevertheless, this study only focuses on the aggregated characteristics of each region (as described by MFDs) without integrating the network structure and topology. This is a necessary step towards implementability of the DBL in reality. It is also interesting to investigate if the network performance (e.g. MFD shapes) are sensitive to the exact choice of dedicated lanes or only the aggregated fraction of DBL compared to total network size matters. This is an interesting conjecture to further investigate. Geroliminis *et al.* (2014) and Chiabaut (2015) conclude that the classical MFD does not account for all interactions in multi-modal traffic, and that traffic flow depends on the mix of cars and buses.

Building on the study of the macroscopic approach of Zheng and Geroliminis (2013), the objective of this work is to analyze the problem in a microscopic level and propose a methodology for finding the best set of roads in an urban network, that should have an DBL, by integrating a detailed traffic model into a simple optimization framework with the objective of minimizing the total travel time for passengers. Despite the detailed characteristics of this model, its formulation (with mainly binary variables) allows its integration in efficient optimization techniques. In order to achieve this, the Store-and-Forward (SaF) queueing model, which has been used as a simulation model in several traffic related studies, e.g. in Aboudolas *et al.* (2009) and Kouvelas *et al.* (2014), is utilized as a building block in the present study, due to its dynamic attributes and simple structure that enable us to study the evolution of congestion and the patterns of spillback propagation for any possible DBL layout with a relatively low computational cost.

The process of examining the impact of different DBL schemes in the performance of the traffic network is complex. Any modification of the infrastructure characteristics (e.g. change of the DBL plan) can influence the travel behaviour of commuters, i.e. the choice of transport mode (car or bus) and selected route. The network performance for the proposed DBL scheme can be safely estimated only when equilibrium in travel choices has been reached. However, modelling the changes in travel behaviour can be really hard in the framework of an optimization problem where many different DBL plans have to be evaluated and involve highly non-linear functions. In the present study, the focus will be on improving the generation and evaluation of candidate DBL schemes in the microscopic level in order to improve passenger mobility, while the related changes in mode and route patterns will be considered as constant.

The contribution of this study is twofold. First, we describe two extensions of the SaF model, in order to consider more accurately the effect of traffic signals and queueing in the performance of the network. The first extension is related with the dynamic characteristics of traffic signal phasing. The existing feature of SaF model assumes that the capacity of a movement at every intersection is static, estimated by the saturation flow and the green ratio of the phase associated with this movement during the whole duration (without the max vs. zero capacity alternation during traffic signal process). Nevertheless, it has been shown that this alternation can significantly influence the network capacity and the evolution of link queues and spillbacks, especially for networks with closely linked intersections, which is the case in most city centers. The interested reader is suggested to refer to explanation based on variational theory and method of cuts in some of the following papers relating MFD performance and network structure (Daganzo and Geroliminis (2008), Boyac and Geroliminis (2011), Leclercq and Geroliminis (2013)). The second extension is related to the fact that some specific lanes change their usage from mixed to DBL or car-only. While both extensions are simple from a mathematical point of view, they are necessary to create a model which is consistent with the physical characteristics of congestion dynamics at the network level. The second and more substantial contribution is the development of a link-level optimization framework for the DBL allocation problem which is mathematically formulated by integrating the revised SaF model.

The rest of this paper is organized as follows: Next section presents the main methodology of the work. It consists of the description of the SaF model and its extensions and the integration to the optimization framework. The objective is to establish a DBL network that will minimize the passenger travel cost in terms of delay for all passengers and modes of transport. Basic Variable Neighborhood Search (VNS) is used to find better solutions based on a given initial one which represents the state-of-practice solution. A local search algorithm is also created to improve an existing solution based on the nature of the problem. The next section presents the application of the methodology in a large scale CBD area which has been calibrated in previous studies based on real data. Various performance measures and system configurations are presented.

2 Methodology

2.1 Problem Description

The problem that is adressed in this paper can be described as follows: Consider an urban traffic network facing high levels of congestion during peak hours. The topological and traffic control characteristics of the network (e.g. traffic signal plans) are considered known. There exist two modes of transport using this network: buses and private cars. Without loss of generality, the operational characteristics of the bus system (routes and frequencies) as well as the average passenger occupancy in buses for all bus lines and for all links are known. No element of the bus system operations is changed in the context of this work, as the focus is on the allocation of the DBLs. A time-dependent O-D demand matrix feeds the network with private car flow. Assuming that, for the purpose of improving mobility in the network, a macroscopic analysis of the wider region dictates that a specific fraction of road space should be given to DBLs (e.g. by following Zheng and Geroliminis (2013)), we seek to decide on the best DBL plan that will lead to best system performance, which is translated here as minimum total Passenger Hours Travelled (PHT). In other words, a set of links of the network must be selected to have one DBL (the right-most lane of the road becomes a bus-only lane so that buses can easily stop at bus stops to board and alight passengers). It should be noted that DBL assignment is fixed to the selected links during the whole simulation time, which can be representative of the morning or evening peak.

2.2 The extended Store-and-Forward model

The following subsection provides a description of the extended SaF model by introducing the appropriate notation. The SaF modelling technique, first proposed by Gazis D. C. (1963), is utilized to simulate the evolution of traffic inside a network based on queueing theory principles. It is composed of a mathematical structure through which the transfer of flow from upstream to downstream link is determined by taking into consideration the current state of the links in terms of vehicle occupancy. Building on this principle, the traffic state of the network is depicted in terms of the number of vehicles inside each link at every time step of the simulation, which is calculated based on a discrete-time form of a flow conservation equation.

Consider a network that is represented as a directed graph consisting of a set of links Z and a set of nodes N. Every link z is defined by a starting node S_z and an ending node E_z, i.e.

$z = \{S_z, E_z\}, \forall z \in Z$, where $S_z, E_z \in N$. Traffic moves in the direction of the link, i.e. from the starting node to the ending node. Every node n is characterized by a set of incoming links I_n and a set of outgoing links O_n. All links $z \in I_n$ have node n as ending node while all links $w \in O_n$ have node n as starting node. Note that a single node n can be referred as either starting or ending node of different links that are connected to it. Every link z belongs to exactly two sets: the set of outgoing links of its starting node and the set of incoming links of its ending node, i.e. $z \in O_{S_z}$ and $z \in I_{E_z}$. The existing version of the SaF model that is found in the literature (see Kouvelas *et al.* (2014), Aboudolas *et al.* (2009)) is based on the following discrete-time flow conservation equation:

$$x_z(k+1) = x_z(k) + T\left[q_z(k) + d_z(k) - s_z(k) - u_z(k)\right], \quad \forall z \in Z, \tag{1}$$

where $x_z(k)$ denotes the number of vehicles inside link z at time step k (also referred as the "queue" in z for simplicity), $k = 1, 2, ..., K$ is the discrete time step index and T the time step duration (KT is the total simulation time), $q_z(k)$ and $u_z(k)$ represent the inflow and outflow (veh/h) of link z at time step k from all upstream and to all downstream links respectively, while $d_z(k)$ and $s_z(k)$ represent the flow of vehicles (veh/h) that start and finish their trips inside link z at time step k respectively. The demand for trips $d_z(k)$ and the uncongested state of the system in the beginning of the simulation, i.e. $x_z(0), \forall z \in Z$, are exogenously given and considered known. The terms $q_z(k)$, $u_z(k)$ and $s_z(k)$ are defined as follows:

$$q_z(k) = \sum_{\forall i \in I_{S_z}} u_{iz}(k), \quad \forall z \in Z, \tag{2}$$

$$u_z(k) = \sum_{\forall i \in O_{E_z}} u_{zi}(k), \quad \forall z \in Z, \tag{3}$$

$$s_z(k) = t_{z0}(k) q_z(k), \quad \forall z \in Z, \tag{4}$$

In Eq. 2 and 3, $u_{zw}(k)$ denotes the flow of vehicles that is transferred in the specific approach z-w, i.e. from upstream link z to downstream link w, at time step k. Eq. 2 states that the total inflow of link z at every time step k is defined as the sum of the flows leaving all upstream links

-i.e. all links having S_z as their ending node - and arriving at link z. The total outflow of link z is calculated in Eq. 3 in the same logic. In Eq.4, $s_z(k)$ is the link exit flow and $t_{z_0}(k)$ is the exit rate of link z at time step k; the latter defines the percentage of cars finishing their trips inside link z as a function of the flow entering the link. This rate can be based on statistical data or can be provided exogenously by a detailed microsimulator (e.g. Aimsun), as in the present case.

The core assumptions of this model regarding the transferred flow $u_{zw}(k)$ in every approach z-w at time step k are the following: i) it is zero if for the current time step k the receiving link w is full and spillbacks occur, ii) it is equal to the fraction of the queue in the upstream link z that heads to w at time step k but cannot be greater than the saturation flow of the approach z-w if the receiving link w is not full, as in Cell Transmission Model (see Daganzo (1994)). Therefore, the outflow $u_{zw}(k)$ is governed by the following equations:

$$u_{zw}(k) = \begin{cases} 0, & \text{if } x_w(k) \geq \alpha c_w \\ \min\left[S_{zw}(k)\frac{g_{zw}}{C_{E_z}}, \frac{x_z(k)t_{zw}(k)}{T}\right], & \text{else} \end{cases} \tag{5}$$

$$c_z = \frac{l_z L_z}{L_{\text{veh}}} \tag{6}$$

$$S_{zw}(K) = \min\left[S_z t_{zw}(k), S_w\right] \tag{7}$$

$$S_z = 1800 \times l_z \quad \text{(veh/h)}, \tag{8}$$

In Eq. 5, c_w denotes the storage capacity of the receiving link w (veh), i.e. the maximum length of the queue inside this link, S_{zw} denotes the maximum flow that can be transferred from z to w (veh/h), g_{zw} is the green time of the approach, C_{E_z} is the cycle duration of the signal plan in the node E_z that connects the two links and $t_{zw}(k)$ is the fraction of vehicles currently in z that will move to link w after leaving z. The value of this ratio, which is also called turn rate, is related to the paths that drivers follow and can be based on traffic measurements or provided dynamically by a detailed microsimulator which integrates microscopic traffic flow models (e.g. car following) and detailed route assignment models. The parameter α is used to prevent link overloading in cases close to capacity and its value can be set in relation to the chosen T (e.g.

0.95). The storage capacity of every link z is calculated by the Eq. 6 where l_z is the number of lanes of link z, L_z is the length of the link z and l_{veh} is the average vehicle length. In Eq. 7, $S_{zw}(k)$ is defined as the minimum of two terms: the saturation flow S_z of the sending link z multiplied by the turning rate $t_{zw}(k)$ and the saturation flow of the receiving link w, S_w; saturation flows are calculated by Eq. 8.

It should be noted that the paths of vehicles circulating in the network are not known to the model but the impact of route choice in the propagation of congestion is expressed through the values of the turning rates $t_{zw}(k)$, which are considered known for all pairs of links z and w, for which $E_z \equiv S_w$ at every time step k. These values can either be constant for the whole simulation time or vary between time steps in the same way as traffic trends vary throughout the day. In case that a specific movement between two links z and w is forbidden, $t_{zw} = 0$.

The extension that was made in the existing version of the SaF model concerns the calculation of the flows u_{zw}. In order to have a more detailed simulation of flows by considering the effect of traffic lights, the ratio of green time over the cycle length that appears in Eq. 5 is removed from the formulation. The outflow $u_{zw}(k)$ is calculated as the product of the value produced by the branch function of Eq. 5 and a binary function $\eta_{zw}(k)$, which takes the value one if the approach z-w has right-of-way at time step k based on the detailed traffic signal plan and is zero otherwise. The binary function $\eta_{zw}(k)$ is defined as follows:

$$\eta_{zw}(k) = \begin{cases} 1 & \text{if z-w has ROW at time step k} \\ 0 & \text{else} \end{cases} \tag{9}$$

Then the outflow is calculated by the following equation:

$$u_{zw}(k) = \eta_{zw}(k) \times \begin{cases} 0, & \text{if } x_w(k) \geq \alpha c_w \\ \min\left[S_{zw}(k), \frac{x_z(k) t_{zw}(k)}{T}\right], & \text{else} \end{cases} \tag{10}$$

The SaF modelling scheme is utilized in the present work because of its ability to dynamically simulate the evolution of traffic during high congestion periods; due to these characteristics the model can capture the pattern of spillback propagation and estimate the delays in a more realistic way compared to steady state models.

2.3 Modelling of the network

In order to simplify the simulation process, the network in this work consists of four sets of links: i)Origin links [O], ii)Intermediate links [I] and iii)Destination links [D] iv) Virtual queues [VQ]. Virtual queues are virtual links with infinite storage capacity, placed directly before origin links. Travel demand is generated only in virtual queues and is transferred to origin links in the next time step on condition that there is available space, in accordance with Eq. 5. In case origin links are full, the newly generated traffic demand becomes part of the virtual queue which discharges based on the traffic conditions in origin links. An origin link is connected upstream only with its virtual queue, from which it receives inflow. Trips end only in destination links, which act like "sinks" in this model. This means that they only receive traffic from their upstream links having infinite storage capacity, while their state is not modelled, as they are always considered as not congested. Intermediate links are placed between origin and destination links. Based in this modelling structure, Eq. 1 for origin and intermediate links is simplified as follows:

$$x_z(k+1) = x_z(k) + T\left[q_z(k) - u_z(k)\right], \quad \forall z \in O \cup I, \tag{11}$$

For virtual queues, Eq. 1 and Eq. 5 are written as:

$$x_{VQz}(k+1) = x_{VQz}(k) + T\left[d_z(k) - u_{VQz}(k)\right], \quad \forall z \in O, \tag{12}$$

$$u_{VQz}(k) = \begin{cases} 0, & \text{if } x_z(k) \geq \alpha c_z \\ \min\left[S_z(k), \frac{x_{VQz}(k)}{T}\right], & \text{else} \end{cases}, \quad \forall z \in O \tag{13}$$

In Eq. 12 $x_{VQz}(k)$ denotes the number of vehicles that are in the virtual queue upstream of origin link z at time step k and $u_{VQz}(k)$ is the flow entering origin link z at time step k which is calculated in Eq. 13 in the same logic of Eq. 5.

2.4 Optimization Framework

The objective of the present study is to identify the optimal road set, in terms of total passenger delay, for the installation of dedicated bus lanes in a network with given topological characteristics and traffic control plans, where a bus system with predefined routes and frequencies exists. In order to mathematically formulate the problem, we define a set of binary variables $Y = \{y_z | \forall z \in Z\}$, where y_z indicates the setting or not of an DBL in link z, as follows:

$$y_z = \begin{cases} 1, & \text{DBL in link } z \\ 0, & \text{else} \end{cases}, \qquad \forall z \in Z \tag{14}$$

In most real networks, not all links are considered for bus lane allocation because of limited space, absence or sparsity of bus routes or other potential constraints. In this case a subset of all links $Z_f \subset Z$ is defined a priori and the set of decision variables of our problem is $Y = \{y_z | z \in Z_f\}$, while for the rest of the links, $y_z = 0$. The dedication of space to exclusive bus use leads to a decrease in the storage capacity and the saturation flow of the link. This change is modelled by modifying Eq. 6 and 8 respectively, as shown below:

$$c_z = \frac{(l_z - y_z)L_z}{L_{\text{veh}}}, \qquad \forall z \in Z_f \tag{15}$$

$$S_z = 1800 \times (l_z - y_z) \quad \text{(veh/h)}, \qquad \forall z \in Z_f \tag{16}$$

It is assumed that when $y_z = 1$, link z has one dedicated lane, which is the right-most lane of the link (left in some countries) to facilitate boarding of passengers in bus stops. Our goal is to establish a DBL network that will minimize the passenger travel cost in terms of total passenger delay. Therefore, for the purpose of comparison among different bus lane allocation schemes, an estimate for the total travel time for all passengers (Passenger Hours Travelled - PHT) for the whole simulation time is given based on the evolution of queues $x_z(k)$ inside all links as:

$$PHT = \sum_z \sum_k \left[x_z(k)\xi + \sum_l \left[(1 - y_z)P_l(z,k)\left(1 + \frac{x_z(k)}{c_z}D\right) + y_z P_l(z,k) \right] \right] T \tag{17}$$

In the above equation, the first term refers to the total travel time of passengers in private cars, where ξ is the average private vehicle occupancy (passengers). The second term refers to the total travel time of passengers travelling by bus, where l is the index of the bus line and $P_l(z,k)$ is the number of passengers travelling on-board line l buses through link z at time step k (pass/hour). This value can be obtained from statistical data provided by the bus company as $P_l(z,k) = f_l(k)P_l^b(z,k)$, where $f_l(z)$ is the frequency of line l (buses/hour) and $P_l^b(z,k)$ is the average number of passengers per bus of line l when traversing link z at time k. The values $P_l(z,k)$ are considered as known inputs in the context of this work. Last, the term $\left(1 + \frac{x_z(k)}{c_z}D\right)$ serves as a scaling factor that linearly "increases" the bus travel time that the model will consider in link z as a function of the link occupancy at the specific time step. This means that in a gridlocked link z, i.e. $x_z(k) = c_z$, the model will count that buses in mixed traffic lanes will spend $(1 + D)$ times the time that buses travelling in DBLs will spend in the same link z. The value of parameter D can be user-defined after trial-and-error simulation experiments and depends on the value of T.

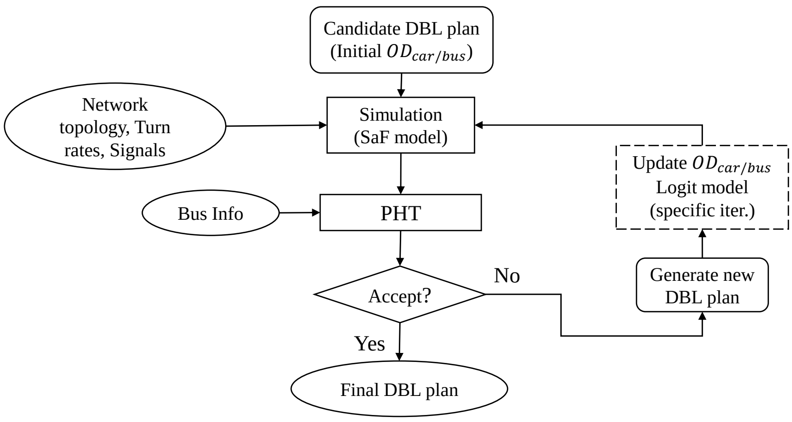

Figure 1: Flow chart of the iterative optimization process. The generation of new solution can be done via various techniques.

It should be noted that the absolute values of PHT given by Eq. 17 do not represent realistic values of trip durations in cases of very low traffic congestion. This is owed to the queueing nature of SaF model, which estimates the travel time based on the service time of vehicles waiting in queue inside each link ($x_z(k)$) but it ignores the actual time that vehicles need to move between consecutive queues (intersections). In order to remove this weakness of the model, proper modification is required, which is still work in progress. However, the metric PHT achieves to evaluate the quality of a candidate solution in comparison to other candidate solutions without being very mathematically and computationally complex.

In Fig. 1 the process of optimizing the DBL network is displayed. The SaF model is used to perform traffic simulation and calculate the value of the objective function. This scheme is quite general and can be adapted to accurately fit any optimization procedure that is based on consecutive steps of evaluation-modification of a current solution. As mentioned in the introduction, any modification of the infrastructure of the traffic network can cause changes in commuters' decisions, such as in mode and route selection. This means that in an iterative optimization procedure as the one in Fig/ 1, the choices of passenger between car-bus and the routes of cars (i.e. the turn rates and the O-D demand matrix for cars) would need to be updated as shown in the dotted line box, until an equilibrium is reached for the specific DBL plan. However, this is not an easy task to perform and depending on the utilized models, it can cause significant increase in the computational cost. In this phase, we do not consider any changes in the passenger decisions, which means that we do not take into account the potential mode shift from cars to buses (or vice versa) or the changing in route choices that the setting of a DBL plan may cause. Here we focus on identifying a promising optimization technique while this issue can be studied in a later phase.

2.5 Algorithms tested

The optimization problem that was described in the previous section is a non-linear combinatorial optimization problem with a finite set of feasible solutions. However, the size of the problem makes it impossible to solve through enumeration because this will require an extremely high computational cost. State-of-practice solutions in this problem would include the analysis and evaluation of a few DBL allocation strategies based on the characteristics of the specific network and of the bus system, e.g. choosing roads with high bus circulation or high demand in bus trips. These solutions, based on practical intuition or experience, could provide some solutions with relatively low value of the objective function (total travel delay). In the current stage of this work, a basic Variable Neighborhood Search (VNS) algorithm is applied to a set of state-of-practice initial solutions. Two neighborhood structures have been embedded in the algorithm. A local search algorithm with a problem-specific structure was also created and applied in the same way.

2.5.1 Variable Neighborhood Search

Variable neighborhood search (VNS) is a metaheuristic for solving combinatorial and global optimization problems proposed by Mladenović and Hansen (1997). The basic idea of VNS lies upon "a systematic change of neighborhood both in a descent phase to find a local optimum and

```
Function BVNS(x, k_max, t_max)
t ← 0
while t < t_max do
    k ← 1
    repeat
        x′ ← Shake(x,k)                        // Shaking
        x″ ← BestImprovement(x′)               // Local search
        x,k ← NeighborhoodChange(x,x″,k)       // Change neighborhood
    until k = k_max
    t ← CpuTime()
return x
```

Figure 2: Basic VNS Algorithm (Hansen *et al.* (2019)).

in a perturbation phase to get out of the corresponding valley" (Hansen *et al.* (2019)). In the latter reference, a thorough description of the method and its multiple variations is provided to interested readers. For the scope of this work, basic VNS algorithm (BVNS) is applied to the DBL allocation problem, in a broader research attempt over different metaheuristic approaches that exist in the literature. Our objective is to identify a promising method that can prove efficient by relating to the problem's nature.

The algorithm that we used is presented in "Algorithm 7" in Hansen *et al.* (2019), which is given in Fig. 2. The focus of our research lies upon defining a set of effective neighborhood structures for VNS algorithm that would intuitively relate to the specific problem. In this work we test the VNS algorithm with two neighborhood stuctures:

- Based on a current solution, randomly pick two links with a DBL and two links without DBL and swap them. The total number of links with DBL remain constant.
- Based on a current solution, repeat the process of the previous neighborhood structure but the selection will not be random but will be based on a probability distribution, where links that improve PHT if a DBL is set on them (assuming all other links in the network have no DBL) have higher probability to be chosen for DBL addition and lower probability to be chosen for DBL removal.

2.5.2 Local Search Algorithm

A local search algorithm based on physical intuition has been designed and tested, with the purpose of providing a significantly improved solution by starting from an initial, state-of-

```
Require: Y = {y_z | ∀z ∈ Z_f}
  obj_0 ← PHT(Y)
  impr ← obj_0
  while impr > 0 do
    i ← 1
    for all y_z = 1 do
      y_z ← 0
      Obj(i) ← PHT(Y)
      Ind(i) ← z
      y_z ← 1
      i = i + 1
    end for
    i_min ← index(min(Obj))
    L1 ← Ind(i_min)
    i ← 1
    for all y_z = 0 do
      y_z ← 1
      Obj(i) ← PHT(Y)
      Ind(i) ← z
      y_z ← 0
      i = i + 1
    end for
    i_min ← index(min(Obj))
    L2 ← Ind(i_min)
    y_L1 ← 0
    y_L2 ← 1
    obj_1 ← PHT(Y)
    impr ← obj_0 − obj_1
    obj_0 ← obj_1
  end while
```

Figure 3: Schematic description of the optimization process for DBL allocation.

practice good quality solution with relatively short computational cost. The algorithm builds on the logic of the steepest decent algorithm (SD), which has been established for continuous variables (as it involves an estimation of derivatives). The SD method approaches the minimum in a "zig-zag" manner, where the new search direction is orthogonal to the previous. The direction chosen is the one where the objective function decreases faster. The search starts at an arbitrary point and then slides down the gradient, until we are close enough to the solution. Given that the decision variables are binary (make a lane DBL or not) the change in the objective function is estimated by adding or removing a single DBL from the previous step. The use of the SD method starting from a random solution could be trapped at a local minimum that might be far from the optimal solution. Thus, the initial solutions based on the state-of-practice provide some "good" solutions as starting point and based on the fact that we start the SD from different points, there is an expectation that we get closer to the global optimum. The function of this algorithm is displayed in the form of a pseudocode in Fig. 3.

Starting from an initial solution represented by a vector of binary variables, the algorithm at each step removes one DBL of the initial solution from one road and adds one DBL in a road where there was no DBL before. In order to select the two links to make the exchange in the best possible way, the algorithm first evaluates the impact of removing one bus lane from every single link where one exists (i.e. for every link z where $y_z = 1$ in the current solution) while keeping the rest of the solution unchanged and selects the one among them for which the removal of the DBL will result in the lowest (best) value of the objective function. Similarly, the same process is followed for the addition of an DBL (i.e. for every link z where $y_z = 0$ in the current solution) and the second link is identified. Then the values of the decision variables of the two selected links change (0 to 1 and vice versa). The above process is iteratively repeated until the solution cannot further improve. This scheme of modifying the solution maintains the same number of links with a DBL as in the initial solution. However, if our constraint for putting DBLs is expressed by a maximum percentage of total road space, the algorithm should be tested with several initial solutions, which would give approximately the same fraction of road space to DBLs but with different number of links hosting them, in order to better explore the solution space.

3 Case Study

The proposed methodology is tested through simulation experiments for a real urban network with several bus lines. In this work, the Downtown area of San Francisco, in California, USA is used as a case study. The network contains several bus lines that run in high frequencies during the peak hour, allowing for carrying high passenger loads; as a result, the public transportation system can assist in alleviating traffic congestion in the city. A realistic simulation set-up for this multi-modal network is described in the current section.

3.1 Network description

The test site is a 6.5 square kilometer area of Downtown San Francisco (Financial District and South of Market Area), including 156 intersections and 426 links with lengths varying from 70 to 400 meters (see Fig. 4 (a) for the map representation of the area). The number of lanes per link varies from 1 to 6 and the free flow speed is 50 kilometers per hour. In the network there are 92 signalised intersections and 64 where no traffic lights operate. Traffic signals are all

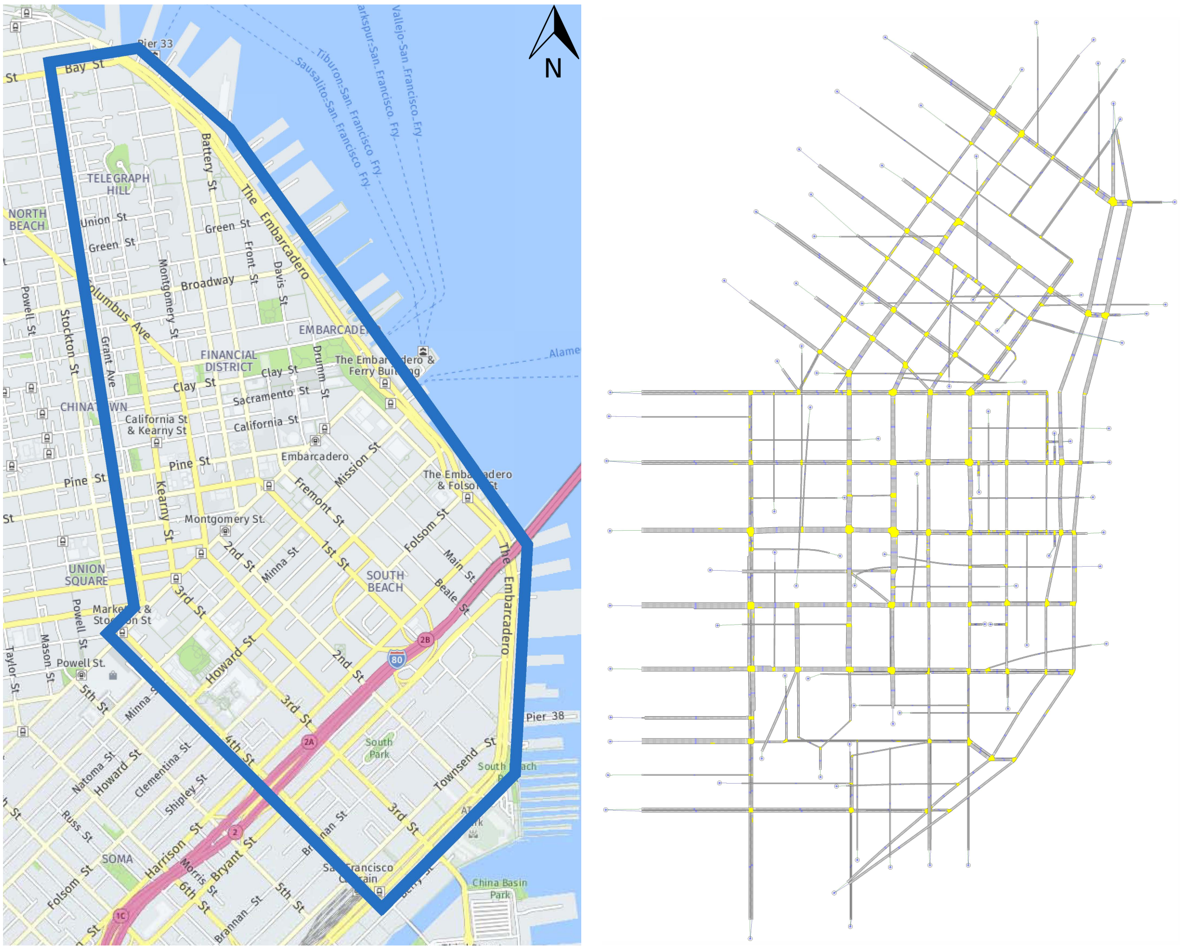

Figure 4: (a) Map of the San Francisco CBD and (b)Microsimulation model of San Francisco CBD in Aimsun.

multiphase fixed-time, operating on a common cycle length of 90 seconds for the west boundary of the area (The Embarcadero), and 60 seconds for the rest of the network.

A well-calibrated microsimulation model exists for this area in the commercial software Aimsun (see Figure 4 (b)). The model includes all the bus lines of the Downtown area of San Francisco, the timetables with the frequencies during the peak hour, as well as a realistic detailed OD demand table that has been calibrated with real data. In the studied area of San Francisco, there are 29 bus lines that run in regular frequencies (e.g. every 10 minutes) and traverse different number of links and bus stops; the information has been provided by the city operators. Out of the 426 urban links, 218 do not serve any bus line, and for the rest 208 the number of associated bus lines varies from 1 to 10 (see Figure 5).

The selecion of the roads that are considered as candidate links is based on space avalability while the current bus routes in the studied network are taken into account. In detail, candidate links are considered links with at least one bus lane and three or more lanes before the setting

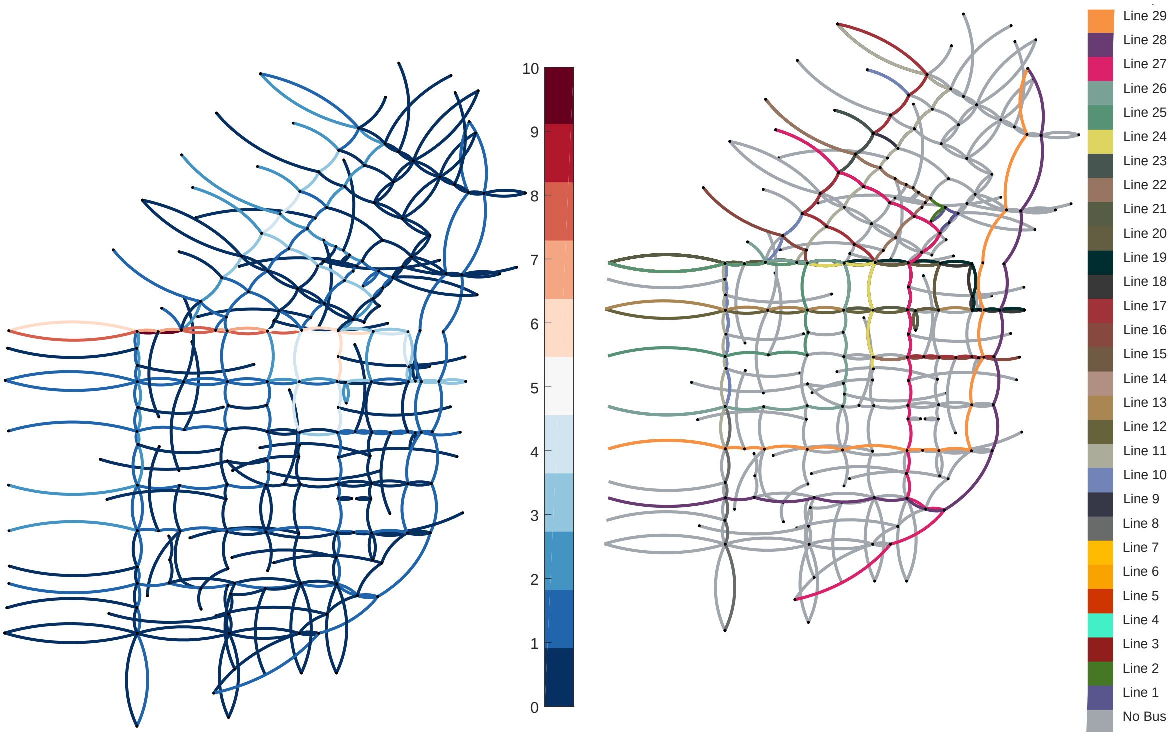

Figure 5: (a) Distribution of number of bus lines in each link; (b) routes of the 29 bus lines of the studied region.

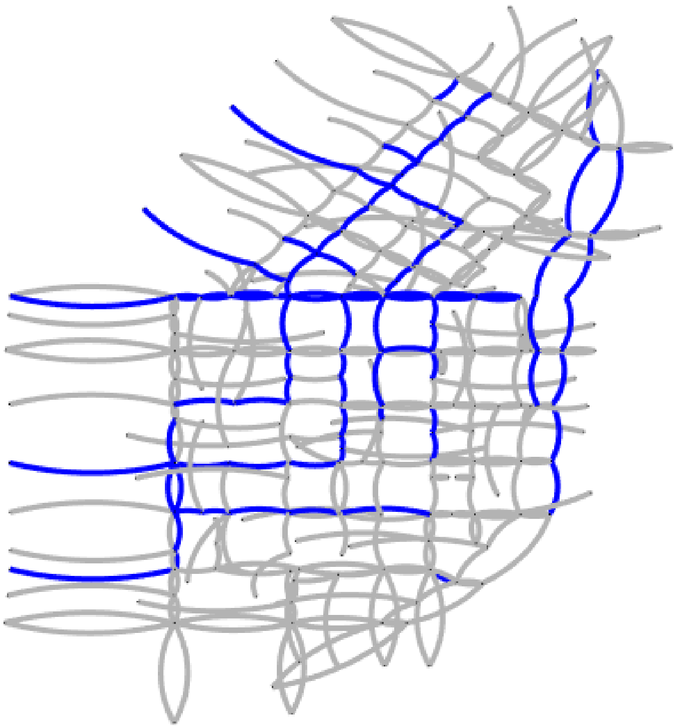

Figure 6: Layout of all candidate links for DBL setting (blue color).

of an DBL. However, as an excpetion we also included some important arterials with high bus circulation. The layout of the candidate links is shown in Fig. 6

For the simulation tests, we assume that in the base scenario there are no designated bus lanes in the network. The research question is to decide the optimal location of the dedicated bus lanes, so as to minimize the total passenger hours travelled (PHT). The simulation step for the microscopic model is set to 0.5 seconds. Furthermore, in order to simulate the adaptivity of drivers and account for route choice effects in the OD demand, the Dynamic Traffic Assignment (DTA) module (C-Logit route choice model) is activated in Aimsun every 3 minutes, a time interval that is consistent with the average trip length in the test area of San Francisco.

3.2 Simulation set-up for store-and-forward model

From the microsimulation network we extract all the data required as input for the store-and-forward model. First of all, we need static data for the topology of the network (i.e. number of lanes, coordinates, connectivity of links), as well as the signal settings per node and per movement (e.g. left, right, through). Moreover, in order to obtain dynamic data from Aimsun (e.g. turning ratios) we need to run some replications and extract the data. In the current work, we use time varying turning ratios (i.e. averages that change every 15 minutes). Another input for store-and-forward that needs to be extracted from Aimsun is the demand in the origins of the network. All these data consist the inputs for the SaF simulation model, which based on the selection of the preferred DBL allocation scheme (i.e. the values of Y_z, also an input for the simulator) evaluated the network performance.

or Sol

4 Results

In the following section, the results of the optimization experiments via VNS and local search algorithms based on the SaF simulation model are presented. The two algorithms take as input an initial state-of-practice solution and return as output an improved solution and the new (decreased) PHT value. Various different solutions have been tested as initial solutions. We present here the results for a set of four of them. These are designed based on intuitive rules related to the problem, as is described below:

- Sol #1: Links with high bus passenger flow
- Sol #2: Links with lots of available road space (high number of lanes)
- Sol #3: Links with high bus frequency and connected

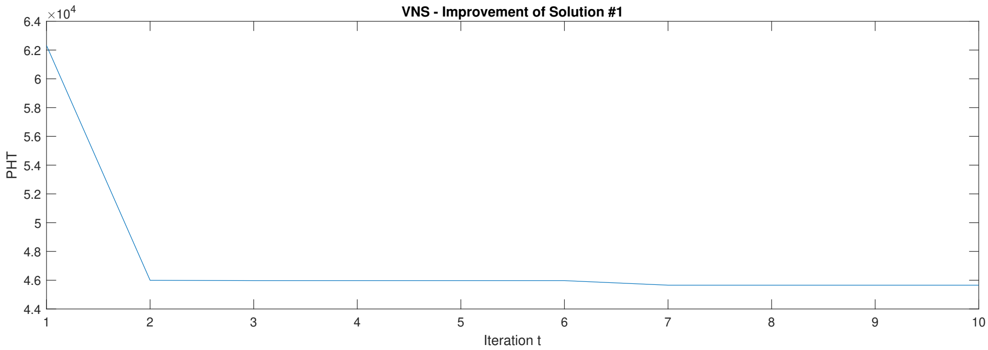

Figure 7: Improvement of Solution # 1 by basic VNS algorithm.

- Sol #4: Random selection (for comparison reasons)

All the above solutions allow the same fraction of the total road space of the network to DBLs which is approximately 3%. While the final solutions have the same number of links with a DBL, the fraction of road space may differ slightly, because links have lengths that vary.

4.1 VNS Results

The basic VNS algorithm was applied for the four initial solutions that were constructed according to the rules described above. Due to high demand in computational time (mainly because of the use of "Best Improvement" algorithm) the construction of the neighborhood for both structures is limited to 7 neighbors. One execution of the algorithm is terminated after 10 iterations (instead of using execution time, we define as t the number of iterations of the *while* loop in 2. In Fig. , we can see the improvement in PHT that is achieved after every iteration t for the four initial solutions. In table 2 we can see the values of PHT of the initial and final solutions produced by basic VNS algorithm. It should be noted that the PHT value of the network with no DBLs is 4.98×10^4. We can see that in highly congested networks, the allocation of DBLs can worsen or improve the situation, which means that proper allocation is crucial for the effectiveness of this transit priority measure. We should note that the improvement shown in Table 2 does not account for a possible mode shift form car to bus, which can improve more the performance of the network.

Table 1: PHT values of initial and final solutions produced by BVNS algorithm

BVNS	Initial PHT $\times 10^4$	Final PHT $\times 10^4$
Sol # 1:	6.23	4.56
Sol # 2:	5.38	4.57
Sol # 3:	6.13	4.72
Sol # 4:	5.29	4.69

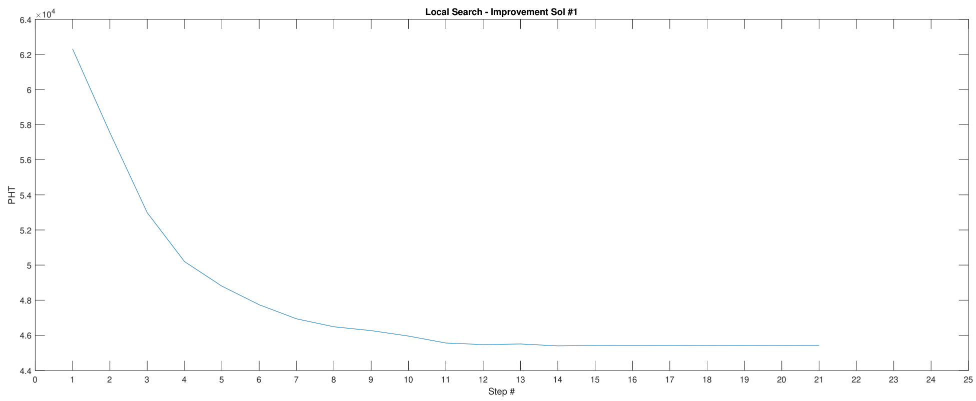

Figure 8: Local Search Algorithm performance: Step by step improvement of initial solution #1.

4.2 Local Search Results

In Fig. 8 we can see the performance of the Local Search algorithm for initial Solution #1. As we can see the algorithm performs well in improving an initial solution in 10 to 15 steps. Please note that each step of the algorithm requires 95 simulation runs (as many as the candidate links for DBL), so to estimate the change in the objective function for adding/removing an DBL from the existing solution. The improvement compared with the best state-of-practice initial solution

Table 2: PHT values of initial and final solutions produced by Local Search

BVNS	Initial PHT $\times 10^4$	Final PHT $\times 10^4$
Sol # 1:	6.23	4.54
Sol # 2:	5.38	4.55
Sol # 3:	6.13	4.55
Sol # 4:	5.29	4.55

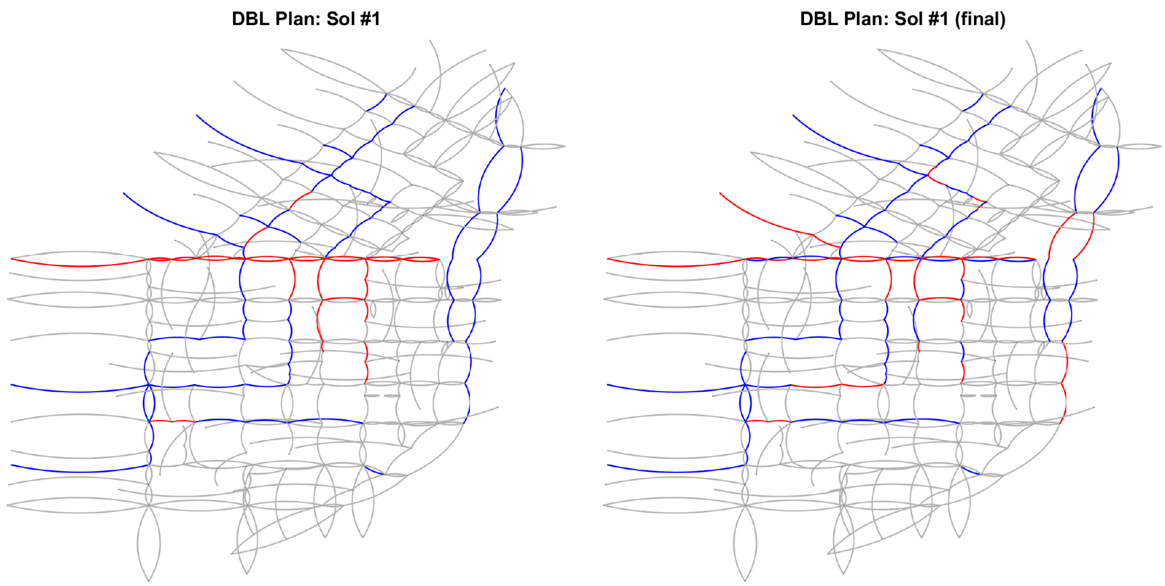

Figure 9: Initial and final DBL allocation scheme for Solution # 1. In blue, all candidate links for an DBL. In red links that actually get an DBL for this solution scheme.

is in the range of 20%, while the state-of-practice solution is about 25% better compared to a random search of 500 replications, showing the importance of carefully designing the location of DBL.

Figure 9 shows the location of DBLs at the beginning and at the end of the optimization algorithm (LS) when starting from initial solution # 1. Please note that the assignment model of the detailed simulator is run once every a few steps to estimate and update the time-dependent turning ratios. Each replication of the SaF model requires only 15-20sec while an assignment in the microsimulation requires 100 times more time. This highlights the importance of integration of the SaF in the solution procedure.

5 Discussion

In this study, the problem of optimal DBS allocation in bi-modal urban networks is addressed via a dynamic modelling approach that is in accordance with the physics of congested networks and a mathematical framework has been designed that is easily integrated in many well known optimization methods existing in the literature. The Store-and-Forward modelling technique is shown efficient in imitating the dynamic propagation of congestion and is in accordance with MFD theory. Two heuristic approaches, basic Variable Neighborhood Search and Local Search have been applied in order to optimize a set of state-of-practice solutions. They both provide promising results as they achieve to improve the given initial solution in a range of

20-25%. Further research is required in terms of techniques exploring the solution space based on problem intuition for the improvement of the heuristics performance and computational time. Moreover, as a secondary future research direction we intend to formulate this complex problem as an MILP and utilize exact methods to solve to optimality.

6 References

Aboudolas, K., M. Papageorgiou and E. Kosmatopoulos (2009) Store-and-forward based methods for the signal control problem in large-scale congested urban road networks, *Transportation Research Part C: Emerging Technologies*, **17** (2) 163–174.

Bingfeng, S., Z. Ming, Y. Xiaobao and G. Ziyou (2017) Bi-level programming model for exclusive bus lanes configuration in multimodal traffic network, *Transportation research procedia*, **25**, 652–663.

Boyac, B. and N. Geroliminis (2011) Estimation of the network capacity for multimodal urban systems, *Procedia-Social and Behavioral Sciences*, **16**, 803–813.

Chiabaut, N. (2015) Evaluation of a multimodal urban arterial: The passenger macroscopic fundamental diagram, *Transportation Research Part B: Methodological*, **81**, 410–420.

Daganzo, C. F. (1994) The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory, *Transportation Research Part B: Methodological*, **28** (4) 269–287.

Daganzo, C. F. and N. Geroliminis (2008) An analytical approximation for the macroscopic fundamental diagram of urban traffic, *Transportation Research Part B: Methodological*, **42** (9) 771–781.

Deng, T. and J. D. Nelson (2011) Recent developments in bus rapid transit: a review of the literature, *Transport Reviews*, **31** (1) 69–96.

Dion, F., H. Rakha and Y. Zhang (2004) Evaluation of potential transit signal priority benefits along a fixed-time signalized arterial, *Journal of transportation engineering*, **130** (3) 294–303.

Eichler, M. and C. F. Daganzo (2006) Bus lanes with intermittent priority: Strategy formulae and an evaluation, *Transportation Research Part B: Methodological*, **40** (9) 731–744.

Gazis D. C., R. B. P. (1963) The oversaturated intersection, *2nd International Symposium on Traffic Theory*, **London 1963**, 221–237.

Geroliminis, N. and C. F. Daganzo (2008) Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings, *Transportation Research Part B: Methodological*, **42** (9) 759–770.

Geroliminis, N., N. Zheng and K. Ampountolas (2014) A three-dimensional macroscopic fundamental diagram for mixed bi-modal urban networks, *Transportation Research Part C: Emerging Technologies*, **42**, 168–181.

Gonzales, E. J. and C. F. Daganzo (2012) Morning commute with competing modes and distributed demand: user equilibrium, system optimum, and pricing, *Transportation Research Part B: Methodological*, **46** (10) 1519–1534.

Guler, S. I. and M. J. Cassidy (2012) Strategies for sharing bottleneck capacity among buses and cars, *Transportation research part B: methodological*, **46** (10) 1334–1345.

Guler, S. I., V. V. Gayah and M. Menendez (2016) Bus priority at signalized intersections with single-lane approaches: A novel pre-signal strategy, *Transportation Research Part C: Emerging Technologies*, **63**, 51–70.

Hansen, P., N. Mladenović, J. Brimberg and J. A. M. Pérez (2019) Variable neighborhood search, in *Handbook of metaheuristics*, 57–97, Springer.

Kouvelas, A., J. Lioris, S. Fayazi and P. Varaiya (2014) Maximum pressure controller for stabilizing queues in signalized arterial networks, *Transportation Research Record: Journal of the Transportation Research Board*, (2421) 133–141.

Leclercq, L. and N. Geroliminis (2013) Estimating mfds in simple networks with route choice, *Transportation Research Part B Methodological*, **57** (EPFL-ARTICLE-188334) 468–484.

Mesbah, M., M. Sarvi and G. Currie (2011) Optimization of transit priority in the transportation network using a genetic algorithm, *IEEE Transactions on Intelligent Transportation Systems*, **12** (3) 908–919.

Mladenović, N. and P. Hansen (1997) Variable neighborhood search, *Computers & operations research*, **24** (11) 1097–1100.

Viegas, J. and B. Lu (2001) Widening the scope for bus priority with intermittent bus lanes, *Transportation Planning and Technology*, **24** (2) 87–110.

Yao, J., F. Shi, Z. Zhou and J. Qin (2012) Combinatorial optimization of exclusive bus lanes and bus frequencies in multi-modal transportation network, *Journal of Transportation Engineering*, **138** (12) 1422–1429.

Yu, B., L. Kong, Y. Sun, B. Yao and Z. Gao (2015) A bi-level programming for bus lane network design, *Transportation Research Part C: Emerging Technologies*, **55**, 310–327.

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