Non-holding-back Dynamic Traffic Assignment with Signal Control: Models and Applications

Tarikul Islam
Swinburne University of Technology
Hawthorn, Victoria, Australia

Supervisors

Prof. Hai L Vu

Dr Antonio Cricenti

A thesis submitted for the degree of

Doctor of Philosophy

March 2018

Abstract

In this thesis, we develop a *linear Non-Holding-Back System Optimal Dynamic Traffic Assignment with Signal Control (NHB DTA-SC)* formulation underpinned by the Cell Transmission Model (CTM). The formulation not only attains System Optimal (SO) solution but also eliminates the unrealistic vehicle Holding-Back (HB) problem. The embedded signal control model in the formulation namely *Signal Control with Realistic Cycle (SCRC)* overcomes the trade-off between signal control cycle-length and cell-length and enables us to strike the right balance between computational complexity and solution accuracy. Note that the cell-length of any CTM based DTA model is directly related to the duration of the discrete time-slot. This discretization time interval has a significant impact on DTA solution accuracy as well as attained vehicle density. As the density contributes to vehicle discharged emission, the above fact motivates us to investigate the effect of time-slot duration on emission. To this end, we propose a *linear Emission-Based Dynamic Traffic Assignment with Signal Control (EB DTA-SC)* formulation. This formulation is free of the HB problem and jointly minimizes emission and travel-time. Our results show that there is a 32% difference between NO_X emission estimated by 60-second and 5-second time-scales.

In addition to combining emission perspective in an NHB DTA-SC framework, we also analyze structures of the optimal solution of such formulation. As the NHB DTA-SC formulation captures real-life traffic propagation characteristics better, it's optimality criterion would provide us the best possible guidelines for traffic management, and modeling signal controllers. In this thesis, we first analyze the NHB DTA problem and present several optimal solution structures of it. We then apply Lagrangian dual decomposition method to decompose an NHB DTA-SC problem and derive a *signal control* subproblem. We also study Karush-Kuhn-Tucker (KKT) optimality conditions of the signal control subproblem and present *locally optimal control structures* for over-saturated, under-saturated, and queue spillback traffic states. The optimal numerical solutions show exact convergence with the introduced theoretical optimal solution structures.

We further extend the idea of the proposed NHB DTA-SC formulation for the public transport use case and present a formulation namely *Bus Priority System Optimal Dynamic Traffic Assignment with Signal Control (BP SO-DTA-SC)*. This framework attains SO solution for both the SC and bus priority in a mixed bus-car traffic environment as well as assures fairness to all the bus and car passengers. We consider Dedicated Bus-only Lanes with Priority (DBLP) for the bus Public Transport (PT). The proposed formulation is linear and takes into account the difference in routing decisions between the buses and cars, i.e., buses have to pass through the predefined locations between depots while cars can freely move around the network before reaching their destinations. Furthermore, we prove that the proposed BP SO-DTA-SC model guarantees a holding-free solution and minimizes the total passenger travel time. The BP SO-DTA-SC is the first model which combines Bus Rapid Transit with Transit Signal Priority in a linear framework.

by Tarikul Islam

Declaration

I hereby declare that I have developed all the contents of this dissertation during my PhD candidature at the Swinburne University of Technology. All the third party materials are cited. This thesis is my original work and has not been submitted to any other university/ institution to obtain a degree.

Tarikul Islam

TARIK

Signature

03/07/18

Date

Dedicated

to my father, mother, and all the siblings

Acknowledgements

First of all, I like to express my deepest gratitude to my principal supervisor Prof. Hai L Vu, and associate supervisors Dr Manoj Panda and Dr Antonio Cricenti for their continuous support and guidance. Throughout my PhD candidature, I have learned many invaluable research and professional skills from them. I believe these skills will support and assist me to become a successful researcher in the upcoming days.

I am grateful to all my friends in Melbourne, especially, Dr Saiful Islam, Dr Samiron Mahmud, Dr Asm Kayes, Dr Tarique Anwar, Dr Arif Moinuddin Siddique, Amr Hosny, and Saeem Khan. I would also like to thank all my cricket-buddies. Completing a PhD is a long and challenging journey, and good friends like them always make things easier.

I am also thankful to Dr Philip Branch and Dr James Hamlyn-Harris for considering me to teach into their course-tutorials. I enjoy working with them. I feel this experience as valuable and believe it will help me to evolve as an academic/network security professional in the future.

I would also like to express gratitude to Nam Hoang, Dr Shiva Pokhrel, Nasrin Akhter, and all other Swinburne colleagues for treating me in a friendly and courteous manner. I wish them lots of success.

Last but not the least, I am indebted to the Swinburne University of Technology for providing me the research fund and required facilities to carry out this PhD study. It has been a pleasant experience studying at Swinburne.

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

The rapid growth in urbanization, where a large number of people migrate to live in cities because of better opportunities and lifestyle prospects, has contributed to the increase in traffic congestion on urban roads. Tackling this increased congestion is a challenging problem around the globe. While expanding road infrastructures requires a long-term commitment and is severely limited by the available space, smart traffic management methods often can bring immediate congestion relief to the cities without a significant investment. As a result, transport planners and traffic engineers are seeking an efficient traffic analysis tool which would enable them to evaluate travel behavior and network performance dynamically within a given short time. After several years of research and development, the *Dynamic Traffic Assignment (DTA)* model has become the most widely accepted modeling option which can be applied to analyze small intersections as well as vast regional traffic networks (Chiu et al. 2011). Moreover, a DTA model would equip us better to deal with not only short-term planning but also to assist us to design more efficient future transport systems. Some examples of the applications of a computationally tractable DTA formulation are: efficient traffic management, intersection modeling, emergency evacuation, major reconstruction planning, study traffic congestion from the perspective of travel time, and reducing vehicle discharged emissions (Janson 1991). Furthermore, DTA would enable us to plan, develop, and to determine policies for a *Bus Priority (BP)* transportation system which is one of the promising strategies proposed to mitigate congestion, vehicle discharged emission, and fuel consumption (Liu et al. 2015a). The above facts motivate us to develop computationally tractable analytical DTA frameworks not only to design and analyze transportation systems but also to seek the global optimal traffic assignment and signal control settings in the existing transport networks. In the next section, we discuss background information about the DTA problems and traffic signal controls.

1.2 Background

DTA is an approach that depicts the interplay between optimal route choice and traffic flows with time and cost constraints. According to the Transportation Research Board (TRB), "DTA describes time-varying network and demand interaction using a behaviorally sound approach" (TRB 2011). In traffic engineering, demands are the number of vehicles leaving an origin within a time window so that they can reach the destination by the desired arrival time. DTA applies to a large problem, where this large problem may consist of several subproblems, and each subproblem corresponds to a different set of decision variables that model the time-varying properties of the transportation system (Peeta and Ziliaskopoulos 2001). DTA has four major elements that are discussed below.

1. Demand: stochastic or deterministic arrival of the vehicles in a network.

2. Dynamic Network Loading (DNL) model: specifies traffic propagation characteristics on the roads.

3. Traffic assignment: departing traffic flows and routes are dynamically determined based on the traffic states. If the objective of the DTA program is set in such a way that it benefits overall system-wide performance, then the attained

solution is called System Optimal (SO) traffic assignment. In contrast, if the objective minimizes individual user's cost, it leads to a User Equilibrium (UE) solution (Boyce and Xiong 2004).

4. Traffic Signal Control (SC): controls traffic flow at the intersections.

In the following subsections, DTA related topics that are relevant to this research are discussed.

1.2.1 DNL Model

Flow propagation model in a DTA framework is called DNL model. For large scale DTA problems, macroscopic DNL models are required. Few popular macroscopic DNL models are Cell Transmission Model (CTM) proposed by Daganzo (1994, 1995), Link Transmission Model (LTM) (Yperman et al. 2006), and METANET (Messmer and Papageorgiou 2000). CTM and LTM are based on first order kinematic wave theory. The LTM is a whole link model and measures number of entering and exiting vehicles only at the start and end of the links. Whereas, CTM discretizes the links into smaller segments called cells and provides more details about traffic propagation (e.g., location of congestion) than LTM. METANET is a second order macroscopic DNL model. Compared to METANET the CTM provides relatively realistic details about queue formation, propagation and dissipation of congestion through kinematic waves and is simpler for urban model based traffic control purposes (Nie 2006). Furthermore, CTM is the most widely used DNL model and covers the full range of fundamental flow-density-speed relationships, which was also validated by field data (Lo 1999). In this thesis, we mainly focus on the CTM based *System Optimal Dynamic Traffic Assignment (SO-DTA)*, because, a SO solution directly reflects system operator's perspective and provides better guidelines for efficient and strategic traffic management than a UE solution. Furthermore, a SO solution attains less Total System-wide Travel Time (TSTT), shorter congestion duration, and lower network

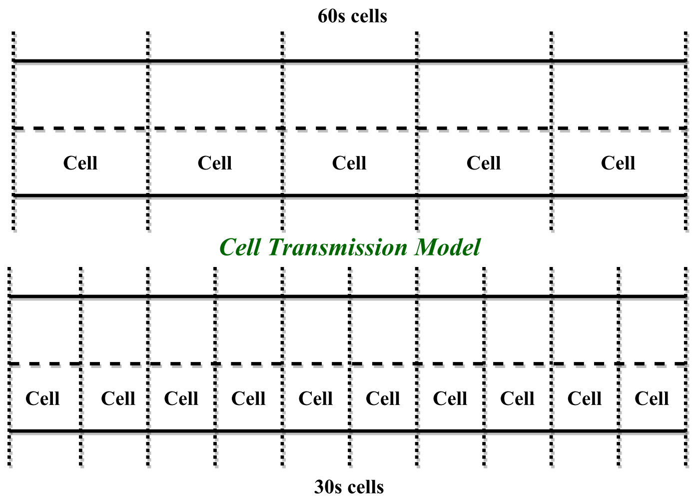


Figure 1.1. Figure shows segregation of a road segment in CTM for the 60s and the 30s discretization time intervals.

cost (Qian, Shen, and Zhang 2012). The original CTM proposed by Daganzo is a nonlinear model. In CTM, the length of each cell $L = v\tau$, where L is the cell-length, v is free-flow speed, and τ is the duration of the discrete time-slot. The free-flow speed is the speed at which drivers operate vehicles on an intersection-free-road under very light traffic condition. The time-slot duration is also called discretization interval. An example of CTM based discretization approach is provided in Figure 1.1 for a 60 seconds (s) and a 30s time-slot duration. In the figure, we see that with shorter time-slot the cell-length decreases. The cell-length also has a significant impact on solution accuracy, i.e., the shorter the cell-length, the higher the accuracy of the DTA solution (Munoz et al. 2003; Sun and Bayen 2008). Note that the nonlinear CTM is computationally expensive and may not find an optimal solution while analyzing large networks.

1.2.2 Vehicle Holding-Back (HB) Problem

One of the traditional approaches in the literature to obtain SO-DTA solution is to formulate a tractable linear optimization problem where the nonlinear CTM terms representing the traffic dynamics in the network are linearized (Ziliaskopoulos 2000).

A definition of linear optimization problem is provided below.

Definition 1.2.1. According to Boyd and Vandenberghe (2004), an optimization formulation is linear if the variables $x, y \in \mathbb{R}^n$ are linearly dependent in the objective function $\mathcal{F}_0$ and in all the constraints $\mathcal{F}_1, ..., \mathcal{F}_m$ such that they satisfy

$$\mathcal{F}_i(\alpha x + \beta y) = \alpha \mathcal{F}_i(x) + \beta \mathcal{F}_i(y),\ i = 0, 1, ..., m.$$

Here $\alpha, \beta \in \mathbb{R}$. If an optimization problem does not satisfy the above condition then that formulation is called a non-linear problem.

A nonlinear equation in a framework leads to a nonlinear formulation which is more complex and such formulation cannot guarantee a feasible solution even for a small network. For realistically large networks, the complexity of the problem tremendously increases, and a nonlinear model may become computationally intractable. Although the linearized model provides a correct optimal objective value, this linearization introduces a new dilemma known as the vehicle HB problem. Along with CTM the HB problem is also evident in other linearized DNL models such as Two Regime Transmission model (TTM) (Balijepalli, Ngoduy, and Watling 2014) and LTM. In the HB SO-DTA solution vehicles are held back in a link even though there is capacity available downstream (Aziz and Ukkusuri 2012a; Carey and Subrahmanian 2000; Zhu and Ukkusuri 2013). It is worth noting that the HB problem yields an unrealistic traffic propagation solution (e.g., occupancy and flow) as in real-life traffic conditions vehicles do not slow-down within the link, unless there is congestion. This fact makes the HB solution unusable for real-life applications and online traffic management (Zhong et al. 2016; Doan and Ukkusuri 2012). A DTA formulation that does not hold back traffic and still obtains a SO solution is called *Non-Holding-Back (NHB) SO-DTA* framework. In the next section, issues related to combining signal controls in the DTA formulations are discussed.

1.2.3 DTA with Traffic Signal Control (SC)

Traffic signal control is one of the key determining factors of the way vehicles traverse a traffic network. As an SC directly regulates the traffic flow at an intersection, it also controls the traffic evolution throughout the whole network. It has also been recognized in the literature that the DTA framework is incomplete without a signal control model (Ukkusuri, Ramadurai, and Patil 2010). There have been several control models proposed in the literature to be combined with SO-DTA formulations. All these control models are mixed-integer or nonlinear formulations and computationally expensive. A linear signal control model requires cycle-length (C_L) has to be strictly equal to the time-slot duration, which is, $C_L = \tau$ (Islam et al. 2017; Ukkusuri, Ramadurai, and Patil 2010). As an example, for a practical cycle-length of $C_L = 60s$, the time-slot duration must has to be $\tau = 60s$. If we assume that the free-flow speed is $v = 60$ kilometers (km) per hour (hr), then for $\tau = 60s$, the cell-length becomes $L = 1$km, which is too long to capture traffic dynamics accurately and compromises the accuracy of the solution. It was discussed before that the shorter the cell-length, the higher the accuracy. As a result, this cycle-length equal to time-slot requirement of a linear signal control model imposes a trade-off between DTA solution accuracy and realistic cycle-length. If an SC is embedded in the SO-DTA formulation, then this framework is termed as *System Optimal Dynamic Traffic Assignment with Signal Control (SO DTA-SC)* formulation. Note that the HB problem also results in unrealistic signal control green proportions. This is because in the HB SO-DTA solution, the source only sends as much traffic into the network as its bottleneck(s) can handle, while most of the surplus demands are held back at the origin, which results in hardly any queues (or congestion) forming at the intersections. Throughout this thesis we call the SO DTA-SC with the Holding-Back problem as *Holding-Back Dynamic Traffic Assignment with Signal Control (HB DTA-SC)* framework. On the other hand, we term the SO DTA-SC formulation that jointly attains *System Optimal* *and Non-Holding-Back (NHB)* solution as NHB DTA-SC framework.

1.2.4 Properties of Optimal Solutions

Traffic dynamics of an NHB DTA-SC formulation closely follow real-life traffic propagation characteristics. This is because in the NHB DTA-SC framework vehicles must move into the downstream link if there is space available. As a result, the properties of the optimal solution of the NHB DTA-SC formulation provide valuable guidelines for traffic management strategies and optimal operating points. Locally optimal structures can be directly plugged-in to the transportation system, e.g., at a junction for minimizing intersection delay. Furthermore, these locally optimal solution structures can be implemented based on the traffic condition (e.g., congested or non-congested) at a particular time of the day (e.g., peak or off-peak hours) and the constrained dynamic assignment problem can be segregated into several static problems (Wie, Tobin, and Friesz 1994). This approach would substantially simplify the problem. Consequently, the computational complexity of the problem would significantly reduce. As a result, it is of great importance to understand the optimal solution structures of the DTA framework. One can analyze the optimal solution of the primal DTA problem or the dual of the DTA formulation. The primal analysis consists of solution attributes such as HB or NHB, and optimality criteria regarding delay associated with traffic flow, occupancy, and free-flow speed (Ziliaskopoulos 2000). A SO-DTA problem is a marginal cost pricing problem (Merchant and Nemhauser 1978b). Therefore, it is possible to apply the Lagrange multiplier method for marginal cost analysis. By using a Lagrangian dual decomposition method, the traffic assignment problem can be decomposed into several subproblems. Then one can analyze the Karush-Kuhn-Tucker (KKT) necessary conditions of the subproblems to determine marginal costs of the Lagrange multipliers and locally optimal structures. These local structures would provide us “off the shelf ready” guidelines to manage real-life traffic networks in a distributed manner. However, the properties of an NHB DTA-SC solution are barely covered in the literature and need to be explored further. DTA can also be applied to reduce vehicle discharged emissions, which is discussed in the following section.

1.2.5 DTA to Reduce Emissions

The objective function of a DTA model can be set to minimize the TSTT or to reduce emissions. When the objective function is defined in such a way that it reduces Total System-wide Emission (TSE), then this formulation is called *Emission Based Dynamic Traffic Assignment with Signal Control (EB DTA-SC)* formulation. Note that minimizing travel time does not necessarily reduce emissions (Zhang, Lv, and Ying 2010). Similarly, lowering emissions may result in very long TSTT (Aziz and Ukkusuri 2012b). Therefore, a formulation which would jointly reduce trip time and emission is highly desirable. Vehicle discharged emissions (e.g., NO_X, CO) are directly related to vehicle speed which can be measured using a macroscopic average speed model (e.g., using discrete DTA model) or by implementing microscopic models such as Virginia Tech microscopic energy and emission model (VT-Micro) presented by Ahn (2002). The macroscopic models consider average speed of a group of vehicles within a DTA time-slot. Whereas, the microscopic models account speed, acceleration, and deceleration of the each of the vehicles individually to estimate emission. In general, microscopic models are nonlinear formulations and impose very high computational complexity. For analyzing complex and significant problems such as traffic networks, a linear and computationally tractable DTA formulation is required. DTA would also equip us better for modelling and analysis of public transit systems, which is discussed in the next section.

1.2.6 DTA to Improve Public Transit

A DTA model also enables us to plan and analyze public transit systems in a multi-class traffic environment. It is well known that improving public transit is one of the most promising strategies to reduce congestion (Wu and Hounsell 1998; Eichler and Daganzo 2006) on the urban roads which would eventually reduce vehicle discharged emissions and fuel consumptions (e.g., NO_X, CO) (Ernst 2005; Satiennam, Fukuda, and Oshima 2006). There are different public transit modes such as high-speed rail, subway, tram, and bus. On the one hand, rail, subway, and tram transit modes require substantial financial investments, overly long construction times, and significant maintenance costs. On the other hand, bus Public Transport (PT), having lower capacity, is slow and unreliable in its current form (Dong et al. 2017). As a result, a smarter public transit system is essential to provide the speed and reliability offered by trains/subways as well as flexibility and economy of the buses. One such solution is Bus Rapid Transit (BRT). BRT is a bus transit mode which integrates Intelligent Transport Systems (ITS) and Dedicated Bus-only Lanes with Priority (DBLP) to provide rail-like reliable, high-speed, and low-cost transportation service (Robert 2013; Deng and Nelson 2011). However, unlike rail or subways, BRT vehicles would compete with other classes of vehicles at intersections if no priority or control scheme is implemented. As a result, the BRT operational efficiency can be substantially improved by implementing Transit Signal Priority (TSP) control schemes (Al-Deek et al. 2017; Zhou, Wang, and Liu 2017). A TSP control policy gives priority to the BRT vehicles (i.e., buses) at the signalized intersections to reduce their intersection delay. Due to these facts, BRT with TSP (BRT-TSP) has been deployed in many cities of Asia, Europe, North America, and Australia. A unified mathematical framework will significantly support transit planners and operators to design and assess a BRT-TSP system, transport policies, and management strategies. DTA framework is widely accepted not only for designing, planning, strategically managing, and determining policies of the BRT-TSP system but also for evaluating important public transit investments (Cheung and Shalaby 2017). In the literature, BRT-TSP is formulated either as mixed-integer (Liu et al. 2015a; Cheung and Shalaby 2017) programs or as non-linear formulations (Li and Ju 2009) that are computationally expensive and cannot be applied to analyze large networks. To formulate a linear BRT-TSP framework, when one relaxes nonlinear traffic flow propagation rules, then the unwanted vehicle HB problem (Ziliaskopoulos 2000; Mesa-Arango and Ukkusuri 2014) arises. Formulating an NHB SO-DTA framework again requires mixed-integer (Doan and Ukkusuri 2012) or nonlinear programming (Shen and Zhang 2014; Smaili, Mammar, and Mammar 2011). As a result, formulating a linear and NHB *Bus Priority System Optimal Dynamic Traffic Assignment (BP SO-DTA)* framework is a challenging research problem. If a traffic signal control is embedded in the BP SO-DTA formulation, then this combined framework is called *BP SO-DTA-SC* formulation. To model TSP in the DTA frameworks, a signal control model that can distinguish between bus and other traffic is required. TSP can be implemented (through an objective function) by maintaining Dedicated Bus-only Lanes (DBLs) (Li and Ju 2009) and using the mixed-integer bus indicator variables at the intersections (He, Head, and Ding 2014). Keeping dedicated bus lanes at the intersections requires a control model that can eliminate intra-phase conflicts. In the Australian phasing convention, an intra-phase conflict would occur between bus and car lanes when the through and left movements are allowed together. This mixed-integer method would impose unnecessary delay to the other vehicle classes due to its on-off nature. Furthermore, it is computationally expensive. It is to mention that all the presented TSP control formulations for DBLP in the literature are highly complex (i.e., mixed-integer or non-linear) and unsuitable for implementing in a DTA framework. In the next section, the research questions related to the topics discussed above are outlined.

1.3 Research Questions and Associated Challenges

In this thesis, we investigate following research questions.

1. How one can formulate a computationally tractable NHB DTA-SC framework?

The major problems associated with this question are discussed below.

- To attain NHB solution, the DTA problem becomes mixed-integer or nonlinear formulation. Furthermore, an NHB solution might not be a SO solution. As a result, formulating a linear NHB DTA-SC framework that would attain SO solutions is a challenging research problem.
- As discussed in Section 1.2.3, a *linear signal control* model can set signal control cycle-length equal to the discrete time-slot used in the DTA framework and imposes a trade-off between accuracy and realistic cycle-length. Formulating a signal control model which is linear as well as can set realistic cycle-lengths without imposing this trade-off is another challenging research problem.

To address these research problems we develop an NHB DTA-SC formulation. This formulation is linear and eliminates the trade-off between accuracy and cycle-length mentioned in Section 1.2.3.

2. The second research question is, “how to add emission perspective in an NHB DTA-SC framework”? The major problems related to this research question are outlined below.

   - In general, emission models are nonlinear. As a result, while analysing large networks by using these models, attaining optimal solutions becomes very difficult. Also, NHB formulations are mixed-integer or nonlinear formulations. As a result, integrating an emission model into an NHB framework and formulating a linear and NHB *Emission Based Dynamic Traffic Assignment with Signal Control (EB DTA-SC)* framework is a cumbersome research task.
   - Emission minimization in the DTA problem may substantially increase travel time which is not desired. Additionally, minimizing travel time

might not reduce emissions. Thus, formulating a unified framework that *jointly minimizes emission and travel time* is also a complex research problem.

We respond to these research problems by formulating a linear NHB EB DTA-SC formulation which jointly minimizes travel time and vehicle discharged emissions.

3. The third research question we investigate is, "what are the optimal solution structures of an NHB DTA-SC formulation and the embedded traffic signal control"? Analyzing properties of the optimal solution of a DTA framework comprises the following major challenges.

   - An NHB SO traffic assignment framework would help us to reduce delay on the roads by maximizing traffic flows. However, theoretically defining properties of traffic propagation in the NHB DTA-SC formulation and mathematically proving them are demanding research tasks.
   - To find optimal signal control structures, one needs to decompose the NHB DTA-SC problem into several subproblems. Then marginal cost analysis has to carry out using KKT necessary conditions and constraint qualifications. Analyzing interactions between marginal costs and impact on global optima to derive locally optimal signal control structures is another challenging research task.

   In response to this question, we analyze properties of the optimal solution of an NHB DTA-SC formulation and deduce several locally optimal theoretical structures of the embedded signal control model in the formulation.

4. The fourth and final question we address is, "how one can formulate the BRT-TSP system as a computationally tractable *BP SO-DTA-SC* framework"? The

major research problems connected with this question are explained below.

- It was discussed in Section 1.2.6 that formulating BRT in a DTA framework requires implementing bus priority and eliminating conflicts at the diverging and merging links, which requires mixed-integer or nonlinear programming. For a large network use-case, mixed-integer and nonlinear models are very difficult to solve and may not even find a feasible solution.
- Similarly, TSP is commonly implemented in a DTA framework by using mixed-integer bus indicator variables. Whereas, a linear and continuous TSP model requires eliminating intra-phase conflicts.

As a result, formulating a linear unified BRT-TSP framework is a tangled research problem. We solve this problem by proposing a novel BP SO-DTA-SC formulation. The formulation is linear and attains NHB solutions.

The contributions of this thesis related to the above research questions and problems are discussed in the next section.

1.4 Contributions

The contributions of this thesis can be categorized as follows:

1. technical contributions:
   (a) introduction of a novel linear NHB DTA-SC formulation;
   (b) theoretically shown that under certain conditions locally optimal control policies in the NHB DTA-SC framework are globally optimal.
2. practical contributions:
   (a) coupling of the above framework with emission modelling;
   (b) extension of the NHB DTA-SC framework to model BRT-TSP systems.

In the following four subsections the contributions are further discussed where each subsection is related to the corresponding research question and associated problems.

1.4.1 The Proposed NHB DTA-SC Formulation

- In this thesis we present a *linear and continuous NHB DTA-SC formulation* which attains *SO* solution.
- The embedded linear signal control model in the NHB DTA-SC formulation can set practically long cycle-length for any short discrete time-slot.
- We numerically show that a shorter DTA time-slot more accurately locates the position of bottlenecks and congestions.

Details about these contributions are provided in Chapter 3.

1.4.2 The Linear EB DTA-SC Formulation

To estimate emission, in Chapter 3 we have proposed

- a *linear EB DTA-SC* formulation. This formulation is HB free and jointly minimizes *emission and travel time.*

We have also developed a post-processing approach to estimate emission from the DTA solution.

1.4.3 The Structures of the Optimal Solutions

The contributions related to the properties of the optimal solution of an NHB DTA-SC formulation are given below. Details about derivation procedures of these structures and supporting numerical solutions are outlined in Chapter 4.

- We prove that the optimal solution of an NHB DTA-SC formulation is not unique. However, for a single path network, this formulation attains a unique solution.

- We further validate that the NHB DTA-SC formulation is a flow maximization problem.
- By combining the above two properties, we show that the formulation also minimizes overall delay inside the network and at the intersections.

The contributions of this thesis related to the *properties of the optimal traffic signal control* settings are outlined below.

- We decompose the NHB DTA-SC problem into three separate sub-problems that are *occupancy minimization, flow maximization, and signal control.*
- We analyze Karush-Kuhn-Tucker (KKT) optimality conditions of the *decomposed signal control subproblem* in different traffic conditions.
- The analysis suggests following *locally optimal* control policy:
    - in an over-saturated traffic condition apply “equisaturation” policy;
    - in an under-saturated condition deploy the “proportional” control policy;
    - and in a queue spillback scenario utilize the available capacity downstream.

1.4.4 The Novel BP SO-DTA-SC Formulation

In the Chapter 5 of this thesis, we propose a new *BP SO-DTA-SC* formulation. The formulation has following properties:

- it is a linear framework and attains NHB solutions;
- it provides priority right of way to the buses and includes a bus route model;
- and does not occur unnecessary delay to the other vehicles.

The embedded control scheme in the proposed BP SO-DTA-SC formulation, namely, *Transit Priority Enabling Signal Control (TPE-SC)* model has following advantages:

- it eliminates intra-phase and inter-phase conflicts while remains linear and enables the DTA to implement TSP at the intersections.

All the contributions discussed in this section are published or under peer reviews and will be published soon.

1.4.5 Publications

All the published and submitted papers associated with this PhD research are enlisted below.

- Islam, T., H. L. Vu, M. Panda, N. H. Hoang, and D. Ngoduy. 2015. "The Accuracy of Cell-Based Dynamic Traffic Assignment: Impact of Signal Control on System Optimality." In International Conference of Hong Kong Society for Transportation Studies (HKSTS), 2015, Dec.

- Islam, T., H. L. Vu, M. Panda, and D. Ngoduy. 2016. "A study of realistic dynamic traffic assignment with signal control, time-scale, and emission." Published in Journal of Intelligent Transportation Systems, 2017.

- Islam, T., H. L. Vu, M. Panda. 2017. "Decomposition of a Non-Holding-Back Dynamic Traffic Assignment Formulation and Structures of Optimal Traffic Signal Control." Under 2^{nd} review in Transportmetrica B: Transport Dynamics, 2017.

- Islam, T., H. L. Vu, N. H. Hoang, A. Cricenti. 2017. "A Linear Bus Rapid Transit with Transit Signal Priority Formulation." Under 3^{rd} review in Transportation Research Part E: Logistics and Transportation Review, 2017.

In the next section, the organization of the rest of this thesis is outlined.

1.5 Organization of the Rest of the Thesis

Chapter 2

This chapter reviews literature relevant to the SO-DTA with/without traffic signal control formulation, properties of the optimal solutions of such formulations, emission based DTA, and public transport assignment with the signal control problem.

Chapter 3

We present the NHB DTA-SC formulation and the EB DTA-SC formulation in this chapter along with the numerical analysis of the optimal solutions of these proposed methods.

Chapter 4

This chapter introduces the structures of optimal solution of an NHB DTA-SC formulation. Numerical solutions related to all the outlined structures are also discussed in this chapter.

Chapter 5

In this chapter we first propose the objective function, cell models, and bus route model of the BP SO-DTA-SC formulation. Then we introduce the TPE-SC model. Finally, we discuss numerical optimal solutions of the combined BP SO-DTA-SC framework.

Chapter 6

In this chapter we summarize this dissertation by discussing key contributions, interesting results, and potential future works.

Chapter 2

Literature Survey

2.1 Introduction

The mathematical formulation of the *Dynamic Traffic Assignment (DTA)* problem was first introduced by Merchant and Nemhauser (1978a,b). Since its inception, many researchers have made several contributions in the field. In this chapter, we review the following DTA topics:

- *System Optimal Dynamic Traffic Assignment (SO-DTA)*
- *Traffic Signal Control (SC)*
- Structures of the optimal solutions
- DTA with emission consideration
- Public bus transit with priority

The literature review related to the above topics are discussed in the subsequent sections.

2.2 System Optimal Dynamic Traffic Assignment

DTA is a mathematical formulation that provides a basis for the optimization of traffic network performance and planning. Furthermore, DTA provides an approach to determine the optimal path and departure time based on the transportation network characteristics and user behavior (e.g., selfish or social). When a DTA problem is formulated from the system operator's point of view, then this formulation is called SO-DTA. Whereas, if the DTA is formulated to benefit individual user's cost (e.g., user travel-time), then this formulation is called User Equilibrium DTA (UE-DTA). DTA has four major components that are: demand, *Dynamic Network Loading (DNL)* model, traffic assignment, and signal control. The DNL model is the heart of a DTA framework and specifies how traffic propagates over a given network in space and time. The *Cell Transmission Model (CTM)* is the most efficient and widely used DNL model. This research work adopts the CTM because compared to other macroscopic DNL models (e.g., METANET, DYNALOAD, MCKW) the CTM provides relatively realistic details about queue formation, propagation, and dissipation of congestion through kinematic waves (Nie and Zhang 2005; Nie 2006). In this literature review, more emphasis is given to the CTM based traffic assignment approaches.

Daganzo (1994) first proposed the CTM. The model is consistent with Hydrodynamic theory and can predict traffic evolution over time and space. The CTM discretizes a link into several homogeneous segments called 'cells'. The CTM covers the full range of flow-density-speed fundamental-diagram and has been validated by field data (Lo 1999). Later, Daganzo (1995) extended the model by introducing three-legged junctions (i.e., diverging and merging cells) to represent more complex networks. However, this model discretizes links into homogeneous cells which is a limiting factor when a link-length is a fraction of the cell-length. To overcome the uniform cell-length limitation of the CTM, a new method namely Semi-Variable Cell Transmission Model (SV-CTM) was proposed by Hu, Wang, and Lu (2010). Variable cell length and cell density were introduced in the SV-CTM. A road segment was divided into two parts: an ordinary section, and an approach section. The cell length of the approach section was increased to accommodate more traffic at intersection cells. In all these papers very small networks were studied and required computational efforts were not reported. It is to mention that all these models are nonlinear formulations and has very high computational complexity.

Sumalee et al. (2011) extended the CTM and proposed a *Stochastic Cell Transmission Model (SCTM)*. This model includes stochastic demand and supply. Depending on the congesting level, the model defines several operating modes (e.g., free flow-congestion). A small network was used to obtain optimal solutions. Latter Tokuda, Kanamori, and Ito (2017) extended the SCTM by introducing route choice for multiple origin-destination networks. Both these two approaches are mixed-integer formulations and have very high computational complexity. Furthermore, none of the above formulations describe how to add signal controls in these complex formulations. To address the computational complexity issue in the DTA framework, several linear programming based approaches were presented in the literature. These contributions are discussed in the following section. It should be noted that this review mainly focuses on the contributions related to the SO traffic assignment formulations.

2.2.1 Linearized CTM for the SO-DTA Formulation

Based on Daganzo's work, Ziliaskopoulos (2000) proposed a linear programming (LP) based approach for the System Optimal Dynamic Traffic Assignment (SO-DTA) problem. In the mentioned approach, the nonlinear terms representing the traffic dynamics in the CTM are relaxed and linearized. Although it provides a correct optimal objective value, the relaxation of nonlinear-terms introduces a new dilemma known as the vehicle Holding-Back (HB) problem. Furthermore, no traffic signal control models are used in this formulation. Throughout this thesis, we call this formulation HB SO-DTA. In the HB formulation, vehicles are held back in a link despite there is capacity available downstream (Carey and Subrahmanian 2000; Zhu and Ukkusuri 2013). In real-life, vehicles will always move forward as long as there is space available in the downstream link. It is worth noting that with the presence of traffic signal control inside the network, the HB problem also yields unrealistic signal control sequences. This is because, in the HB SO-DTA solution the source only sends as much traffic into the network as its bottleneck(s) can handle, while most of the surplus demands are held back at the origin. As a result, no queues (or congestions) grow inside the network. This kind of traffic propagation does not exist in reality which makes this formulation unusable in real-life traffic assignment applications. Similarly, Li, Ziliaskopoulos, and Waller (1999) formulated a linear programming based SO-DTA framework in which departure times and route choices were optimized. Although this paper resolves first in first out (FIFO) violation in the linearized CTM, the HB problem persists. A linear CTM based SO-DTA problem to address Network Design Problem (NDP) was investigated by Waller et al. (2006). When improving the network has budgetary constraints then this type of problem is known as NDP. The uncertainty of traveler demand for the NDP was scrutinized by Waller and Ziliaskopoulos (2001, 2006). The primary objective of these contributions were to find an optimal capacity for the NDP. Both the traffic dynamics and demand uncertainty were considered in these models. However, none of these contributions resolve the vehicle HB problem. Furthermore, they did not include traffic signal control in the formulation. The existing approaches that address this unrealistic vehicle HB problem are reviewed in the following section.

2.2.2 Non-Holding-Back (NHB) SO-DTA Formulation

There have been a number of contributions reported in the literature that address the HB problem. In particular, approaches based on mixed integer formulations were presented by Doan and Ukkusuri (2012) and Zhang, Yin, and Lou (2010), while non-linear (continuous) programming methods were developed by Shen, Nie, and Zhang (2007) and Carey and Subrahmanian (2000). Shen and Zhang (2014) addressed HB problem in a system optimal setting for many to one network and formulated the HB

SO-DTA problem as a minimum Cost Flow Problem (MCP). An MCP minimizes the cost of moving every unit of vehicles. However, this formulation cannot model queue spill-back traffic condition. It is also shown in this paper that for an HB SO solution, there is always an NHB SO solution. All the above approaches present mixed-integer and nonlinear formulations and have very high computational complexity. As a result, they might not attain solutions for realistic size networks. Furthermore, how a signal control can be implemented in the proposed approaches is not discussed in these papers. Zheng and Chiu (2011) proved that the HB SO-DTA formulation could be formulated as the earliest arrival of the vehicles to destination cells. A destination cell is also called a sink cell. Although this CTM based approach resolves vehicle HB problem at the ordinary and merging cells, it cannot address HB issue at the diverging cells; in addition to that, this method lacks a signal control model in the formulation. Details about these cell structures can be found in Ziliaskopoulos (2000).

A penalty based linear formulation to obtain a NHB SO-DTA solution was presented by Zhu and Ukkusuri (2013). However, this framework requires the correct choice of predefined parameters. As a result, one needs to solve the problem and use the results to determine parameters and iterate to find the correct parameters. This raises a recursive parameter selection problem. Without the right parameters, the solution might not be SO, or sometimes the formulation may not even attain a feasible solution. Doan and Ukkusuri (2015) formulated a path based traffic assignment formulation for multi-OD networks. The path based formulation enumerates all the paths for each of the OD pairs. This formulation resolves the HB issue and accurately captures the marginal path cost for the SO traffic assignment problem. However, the formulation is nonlinear and including a signal control in the formulation is a challenging issue. For a SO traffic assignment problem, variable speed limits and ramp metering were introduced by Como, Lovisari, and Savla (2016). Although this model resolves HB problem, integral equations in the formulation make it usable only for

Table 2.1. The discussed CTM based SO-DTA approaches.

Contribution	NHB	Signal Control	Linear
Daganzo (1994, 1995)	✓	✗	✗
Li, Ziliaskopoulos, and Waller (1999)	✗	✗	✓
Ziliaskopoulos (2000)	✗	✗	✓
Carey and Subrahmanian (2000)	✓	✗	✗
Waller and Ziliaskopoulos (2001, 2006)	✗	✗	✓
Waller et al. (2006)	✗	✗	✓
Hu, Wang, and Lu (2010)	✓	✗	✗
Zhang, Yin, and Lou (2010)	✓	✗	✗
Zheng and Chiu (2011)	✓	✗	NA
Sumalee et al. (2011)	✓	✗	✗
Doan and Ukkusuri (2012)	✓	✗	✗
Zhu and Ukkusuri (2013)	✓	✗	✓
Shen, Nie, and Zhang (2007); Shen and Zhang (2014)	✓	✗	✗
Doan and Ukkusuri (2015)	✓	✗	✗
Como, Lovisari, and Savla (2016)	✓	✗	✗
Tokuda, Kanamori, and Ito (2017)	✓	✗	✗

✓ formulation is/includes; ✗ formulation is not/does not include; Not Applicable (NA).

small networks.

Table. 2.1 provides an overall summary of the DTA approaches related to SO-DTA and vehicle HB issue discussed above. We see in the table that most of the NHB SO-DTA contributions discussed in this subsection are mixed-integer or nonlinear formulations that also lack signal control models in their formulations. In the next section, existing contributions related to Dynamic Traffic Assignment with Signal Control (DTA-SC) frameworks are discussed.

2.3 DTA with Traffic Signal Control

Traffic signal control is an essential part of the DTA because it directly influences flow and density of traffic on the roads. Therefore, when studying DTA, a signal control must be included in the formulation. In a DTA framework, when both the traffic assignment and signal control settings are determined from the system operator's point of view, then this integrated formulation is called *System Optimal Dynamic Traffic Assignment with Signal Control (SO DTA-SC)* framework. As the SO DTA-SC minimizes overall network-wide delay, SO DTA-SC solutions would provide us the best possible guideline for the traffic assignment procedure. Several signal control models have been proposed in the literature. Almost all of them are mixed-integer or nonlinear models and have very high computational complexity. These models are discussed below.

Chen and Ben-Akiva (1998) developed a game theory based framework to study dynamic control and traffic assignment interaction. A continuous differential equation based Variational Inequality (VI) signal control formulation is provided in the paper. Nevertheless, this differential equation based formulation implies enormous computational burden and cannot be applied to analyze networks of realistic size. Varia and Dhingra (2004) proposed an approach that combines a signal control in a simulation framework. The simulator implements shortest path algorithm to load the flows dynamically. Then *Genetic Algorithm (GA)* is applied to attain a SO signal timing solution. Note that GA is an optimization technique to solve complex non-linear and discontinuous problems (Teklu, Sumalee, and Watling 2007). Pohlmann and Friedrich (2010) presented an adaptive traffic signal control based on the CTM. In this approach, CTM is used for network loading, and GA is used for optimizing traffic signal control. Lertworawanich, Kuwahara, and Miska (2011) proposed a CTM based multi-objective formulation to address signal control issues in an over-saturated traffic condition and attain solution using GA. A simultaneous traffic assignment and signal optimization formulation was presented by Hajbabaie and Benekohal (2015). In this paper, a GA based meta-heuristic approach is proposed to obtain system optimal solution. Similarly, Liu and Chang (2011); Jones et al. (2013); Ren et al. (2013) introduced GA based signal control optimization formulations to be used in the DTA frameworks. To reduce the total network-wide delay of the vehicles, an artificial bee colony based approach was proposed by Gao et al. (2017). In this model, each outgoing flow is modeled as switching function and includes drivers' behavior. The case studies outlined in these papers are limited to small networks as all the proposed signal control models are nonlinear formulations. As a result, these models are highly complex and can be applied to analyze small networks only.

Several mixed-integer signal control models have been proposed and small networks have been studied in the literature such as the models presented by Lo (1999, 2001); Lo, Chang, and Chan (2001), Lin and Wang (2004), and Wang et al. (2013). The mixed-integer problem arises as each of the discrete time intervals is assigned to a certain phase and traffic belong to all other phases have to wait till the respective phase gets the green time. Beard and Ziliaskopoulos (2006) proposed an approach that combines dynamic route guidance and dynamic signal timing. In the proposed technique, the nonlinear CTM contributed by Daganzo was modified to represent signal control at the intersections. Then a mixed-integer traffic signal control was implemented in the formulation and compared with a pre-timed control. Zhang, Yin, and Lou (2010) presented a methodology for designing robust signal timing plans under demand uncertainty. The primary contribution of this paper is the implementation of traffic signal control in a stochastic programming based DTA formulation. Li (2011) investigated the queue blockage problem among the intersection lane groups under the oversaturated condition. An adapted version of the CTM that splits the diverging cell into two sub-cells was presented in the paper, and a signal control was embedded in the formulation. Aziz and Ukkusuri (2012c) presented a unified optimization framework for traffic assignment and signal control. The signal control model explicitly considers intersection delay and lost time from phase switches. The formulation presented in the contribution includes both the uncertainty and resilient factors. A convex signal control model in a two-regime transmission model scenario was presented by Hoang et al. (2016). The proposed model can capture on-off traffic pattern at the intersections. A CTM based optimization formulation to reduce intersection delay was proposed by Mehrabipour and Hajbabaie (2017). In the formulation, intersection-sink cells were introduced. Their objective function maximizes occupancy at the intersection sink cells. All the above models perform well in an over-saturated traffic condition, but they may attain suboptimal solutions in under-saturated and queue spillback traffic scenarios. This occurs due to the binary nature of the mixed-integer models in allocating green times, i.e., one can only allocate green time equal to whole signal control cycle or zero. Moreover, these mixed-integer models are computationally costly and would become limiting factor while analyzing large networks.

The only linear-continuous signal control model for the SO-DTA framework was proposed by Ukkusuri, Ramadurai, and Patil (2010). In this model, the cycle-length can be set equal to the DTA time-slot duration only. Thus, we term this model *Cycle-length Same as Discrete Time-interval (CSDT)* control scheme. The CSDT control requires a large discrete time-slot (e.g., 60-second) to set a practically large cycle-length (e.g., 60-second). It was discussed earlier that accuracy of the DTA solution increases with shorter time-slot duration (Munoz et al. 2003; Sun and Bayen 2008). Therefore, this large discrete time-slot requirement of the CSDT model imposes a trade-off between accuracy and realistic cycle-length.

All the contributions discussed above are summarized in Table. 2.2. In the table notice that most of the SO DTA-SC approaches presented in the literature are

Table 2.2. Existing Signal Control (SC) models in the literature.

Contributions	**SO-DTA**	**NHB**	**Linear**
Chen and Ben-Akiva (1998)	✗	NA	✗
Lo (1999, 2001)	✓	✓	✗
Lo, Chang, and Chan (2001)	✓	✓	✗
Varia and Dhingra (2004)	✓	✓	✗
Lin and Wang (2004)	✓	✓	✗
Beard and Ziliaskopoulos (2006)	✓	✗	✗
Zhang, Yin, and Lou (2010)	✓	✓	✗
Ukkusuri, Ramadurai, and Patil (2010)	✓	✗	✓
Pohlmann and Friedrich (2010)	✗	NA	✗
Li (2011)	✓	✓	✗
Lertworawanich, Kuwahara, and Miska (2011)	✓	✓	✗
Liu and Chang (2011)	✗	NA	✗
Aziz and Ukkusuri (2012c)	✓	✗	✗
Wang et al. (2013)	✓	✓	✗
Jones et al. (2013); Ren et al. (2013)	✗	NA	✗
Hajbabaie and Benekohal (2015)	✓	✓	✗
Hoang et al. (2016)	✓	✗	✗
Gao et al. (2017)	✓	✓	✗
Mehrabipour and Hajbabaie (2017)	✓	✗	✗

✓ formulation is/includes; ✗ formulation is not/does not include; Not Applicable (NA).

mixed-integer or nonlinear formulations and have very high computational complexity. Furthermore, a linear *Non-Holding-Back SO DTA-SC (NHB DTA-SC)* formulation is not yet presented in the literature. This thesis fills this gap by proposing a unified framework NHB DTA-SC formulation. The proposed formulation is linear and computationally tractable. Furthermore, the formulation attains NHB solution as well as overcomes the trade-off between accuracy and realistic cycle-length. The formulation is presented in Chapter 3. In the next section, papers that analyze properties of the optimal solutions of the dynamic traffic assignment formulations are discussed.

2.4 Optimal Solution Structures

The optimal solution structures of the NHB DTA-SC formulation are useful, particularly, for real-life traffic management and traffic signal modeling. There are two ways to derive properties of the optimal solution of a DTA formulation. One is by analyzing the primal DTA formulation. The other is by forming Lagrangian function of the primal problem and decomposing this function into separate subproblems, then analyzing their Karush–Kuhn–Tucker (KKT) optimality conditions. The Lagrangian function consists of the objective function and all the constraints of a formulation where all the constraints are multiplied with marginal cost variables (i.e., Lagrange multipliers). The KKT conditions are first-order optimality criterion for a solution to become optimal. There have been several Lagrange multipliers based methods proposed in the literature to analyze optimal traffic assignment problems. All these approaches focus on marginal cost analysis of nonlinear formulations. Nonetheless, neither the properties (e.g., optimal flow and occupancy) of primal traffic assignment problem are explored nor the optimal traffic signal control structures are studied in the literature. The existing literature relevant to the DTA solution structures is discussed below.

The structure of the optimal solution of a DTA formulation was first discussed by

Merchant and Nemhauser (1978b). In this paper, the DTA problem was formulated as a nonlinear and non-convex problem. A Lagrange multiplier method was applied to analyze marginal costs associated with the optimal traffic flows. KKT conditions were examined for this proposed DTA problem and structures related to minimum cost solution were discussed. Based on Merchant's work, Wie, Tobin, and Friesz (1994) formulated System Optimal traffic assignment as optimal control problems. Then an augmented Lagrangian method was applied to analyze optimal solution structures related to marginal path cost function of the DTA. However, these two approaches assume that traffic density changes instantaneously throughout the whole link. Apparently, these models cannot be used when traffic density is very light. Side constraints based uncapacitated traffic assignment approaches based on Lagrangian multiplier methods were presented by Larsson and Patriksson (1995, 1996); Shahpar, Aashtiani, and Babazadeh (2008). In these methods, dynamic User Equilibrium (UE) assignment problems are solved with the disaggregated simplicial decomposition algorithm supported by a number of side constraints. These concepts can be applied in estimating travel-time. However, all these approaches have limited ability to capture traffic dynamics (e.g., queue spillback), and none of these works have studied optimal signal control structures associated with different traffic conditions. Bell et al. (1997) had proposed an iterative solution procedure to produce estimates of path flows and route travel-times. Lagrange multipliers were used for a number of stop stations and capacity of the links to formulate a UE traffic assignment problem. Notwithstanding, structures of queue formation (i.e., occupancy) and flow propagation are not analyzed in the paper. Li, Ziliaskopoulos, and Boyce (2002); Li, Waller, and Ziliaskopoulos (2003) formulated the HB SO-DTA formulation proposed by Ziliaskopoulos (2000) as a minimum cost flow problem by using the Lagrangian dual decomposition method. The formulation combines trip distribution and traffic assignment. Nevertheless, no optimal solution structures are discussed in these papers.

Nie, Zhang, and Lee (2004) proposed an augmented Lagrangian multiplier approach to formulate link capacitated UE traffic assignment problem. However, this method uses overly simplified DNL model that cannot capture queue spillback or backward traffic propagation. The properties of the SO traffic assignment solution were analyzed by Chow (2009). This paper invokes a deterministic queuing approach to model traffic assignment behavior. KKT conditions for the proposed formulation are derived for the least marginal cost paths. Nevertheless, in a dynamic traffic assignment setting deterministic queue might not be feasible at all times. Furthermore, this method does not explicitly derive structures of queue formation. Carey and Watling (2012) extended the Cell Transmission Model to differential flow density functions as these functions are convenient for derivation and analysis of system marginal costs. The objective function of the formulation minimizes mean path travel-time. In spite of that, the unrealistic vehicle HB problem persists in the formulation. Pisarski and Canudas-de Wit (2012) proposed a CTM based traffic density distribution formulation. By using dual decomposition method, the nonlinear formulation was decomposed into several subproblems and algorithms were proposed to solve these subproblems without reporting any solution structures. To find the most reliable route, a Lagrangian substitution based approach was presented by Xing and Zhou (2011). Later, Xing and Zhou (2013) extended the formulation by including route choice behavior and stochasticity of the traffic network. A Lagrangian decomposition was used to segregate the problem into subproblems. An algorithm was proposed that reduces duality gap and gradually converges to the system optimal solution. Nonetheless, this approach is based on shortest path algorithm and does not capture traffic dynamics such as queue formation and spillback. A CTM based coordinated traffic signal control model was formulated as mixed-integer program by Timotheou, Panayiotou, and Polycarpou (2015, 2014). Then the problem was decomposed into small subproblems. Alternating direction method of multipliers was adopted to solve the problem in a distributed manner. Although the formulation applies penalty terms to resolve vehicle Holding-Back (HB) problem to a certain extent, the formulation is not entirely free of the HB issue. Furthermore, the authors did not analyze optimal properties of the embedded traffic signal control. Bagloee and Sarvi (2015) formulated UE traffic assignment problem as a capacity constrained problem. KKT optimality conditions were analyzed to develop a heuristic method. This approach assumes that travel-times do not change over time, despite under different traffic scenarios the travel-times always change. Intuitively, the optimal properties of this presented approach might not always represent real-life traffic propagation. Li et al. (2007); Ruan et al. (2016) proposed mathematical frameworks to analyze road and parking capacity in UE settings. Although these two papers apply the Lagrange multiplier method to develop heuristic solution algorithms, they do not explicitly report any optimal solution structures.

All the papers discussed above are outlined in Table. 2.3. In summary, most of these papers analyze least marginal cost paths using Lagrange multiplier methods. Many of them use overly simplified DNL models (e.g., exclude queue formation). As a result, their solution structures cannot be used in real-life applications. Surprisingly, optimal solution structures of traffic signal control are yet to be presented in the literature. Furthermore, solution analysis using a primal DTA formulation is not covered in the literature. To address these research gaps, we propose and analyze the optimal solution structures of a CTM-based linear NHB SO-DTA formulation. We derive several interesting structures for the primal traffic assignment problem. Then we decompose an NHB DTA-SC formulation using Lagrangian method and propose several optimal traffic signal control structures that can be easily implemented in the transportation system. These structures are presented in Chapter 4. In the next section, papers related to traffic assignment and signal control formulations with emission considerations are discussed.

Table 2.3. The literature relevant to the structures of the optimal solution.

Contributions	NHB	SO-DTA	SC	Linear
Merchant and Nemhauser (1978b)	✓	✓	×	×
Wie, Tobin, and Friesz (1994)	✓	✓	×	×
Larsson and Patriksson (1995, 1996)	✓	×	×	×
Bell et al. (1997)	✓	×	×	×
Li, Ziliaskopoulos, and Boyce (2002)	✓	×	×	×
Li, Waller, and Ziliaskopoulos (2003)	✓	×	×	×
Nie, Zhang, and Lee (2004)	✓	×	×	×
Li et al. (2007)	✓	×	×	×
Shahpar, Aashtiani, and Babazadeh (2008)	✓	×	×	×
Chow (2009)	✓	✓	×	×
Xing and Zhou (2011)	✓	✓	×	×
Carey and Watling (2012)	×	✓	×	×
Pisarski and Canudas-de Wit (2012)	✓	×	×	×
Timotheou, Panayiotou, and Polycarpou (2015)	×	✓	✓	×
Bagloee and Sarvi (2015)	✓	✓	×	×
Ruan et al. (2016)	✓	×	×	×

✓ formulation is/includes; × formulation is not/does not include; Signal Control (SC).

2.5 Existing Approaches to Reduce Emission

Emission can be estimated at the post-processing stage after a DTA solution is obtained or by directly formulating an emission-based DTA framework. This thesis focuses on developing the latter in a linear NHB context. Several models have been proposed in the literature to include emission consideration in the vehicular traffic assignment and signal control problem. Most of these formulations are nonlinear and highly complex to find a solution for large networks. These models are discussed below.

An Emission Factor (EF) model was proposed by Mensink, Vlieger, and Nys (2000). This model is based on empirical field data in the city of Antwerp. The model considers road type, vehicle type, traffic density, and duration of the trips. Benedek and Rilett (1998) included environmental cost function in the dynamic traffic assignment problem. In this approach, vehicles are assigned with pollution reduction objective. An emission based trip assignment model was proposed by Sugawara and Niemeier (2002). The proposed model was compared to SO, and UE based traffic assignment solutions. It was found in the paper that the emission based DTA can significantly reduce vehicle discharged pollutants. A semi-empirical model for predicting the effect of changes in traffic flow patterns/density on carbon monoxide concentrations was presented by Dirks et al. (2003). Zegeye et al. (2009) combined METANET macroscopic traffic flow model with a microscopic VT-micro emission model. The objective function proposed in this formulation considers minimizing both the emission and travel-time. Aziz and Ukkusuri (2012b) integrated environmental objective in a System Optimal Dynamic Traffic Assignment (SO-DTA) setting. To implement traffic flow behavior, the CTM was adopted in this approach. Then an objective function was proposed that minimizes both the emission and travel-time. A multiple-objective-based approach to mitigate both the congestion and emission was suggested by Friesz et al. (2013). In this method, congestion pricing and emission is formulated as equilibrium constraints. A dynamic speed limit control based flow propagation model was formulated as Markov decision process by Zhu and Ukkusuri (2014). The model captures stochastic nature of demand and supply. The formulation optimizes traffic flow and emission. A nonlinear energy consumption model was embedded in the CTM based VISTA DTA simulator by Levin, Duell, and Waller (2014). The paper also focuses on the effect of road grades on the vehicle route choice. A CTM based variable message sign for the free-way scenario was presented by Liu et al. (2015b).

Two different objectives were proposed in the paper to simultaneously smooth flow propagation on the freeways and reduce vehicle discharged emission. Ma, Ban, and Szeto (2015) showed that under certain traffic conditions free-flow SO solution exists that minimizes emission as well as travel-time. A macroscopic emission model was implemented in this paper. An eco-system optimal *Emission-Based Dynamic Traffic Assignment (EB DTA)* formulation was presented by Lu et al. (2016). The objective function in the proposed formulation minimizes total system-wide emission. A mixed-integer DNL model was introduced in the paper. Then this mixed-integer program was solved using a Lagrangian relaxation-based algorithm. Wang, Zhang, and Zhang (2016) proposed an approach to simultaneously reduce travel-time and emission at car parking by distributing parking permits. The authors used a bi-level objective function to attain optimal travel-time and emission. Luo et al. (2016) formulated a Model Predictive Control (MCP) framework based on METANET, which considers traffic flow efficiency, emission, and fuel consumption reduction. The model also includes uncertainty and disturbance factors in the traffic network. In the framework, route guidance provided by a traffic manager is considered as a decision variable when certain proportion of road users follow the route guidance. Long et al. (2016) proposed a Link Transmission Model (LTM) based traffic assignment formulation to minimize emission. The proposed objective function in this paper jointly minimizes traffic congestion and emission. The formulation also accomplishes NHB solutions. Pasquale et al. (2017) proposed an approach to reduce travel-time and emission simultaneously. The proposed approach combines ramp signal and route guidance to manage traffic flows on the freeways. A non-linear CTM based emission prediction model was proposed by Sayegh, Connors, and Tate (2017); Sayegh (2017). In this approach, speed is calculated from vehicle flow and occupancy in a cell then coupled into an EF model to estimate emission. Although multiple real data sets were used to calibrate the CTM model parameters, small links were tested using the proposed approach. Ma, Ban, and Szeto (2017) formulated a bi-level program to optimize both the travel-time and emission. At the upper level, the objective is to minimize total system-wide travel-time, whereas, the lower level goal is to reduce emissions. Nevertheless, all of the contributions discussed above are mixed integer or nonlinear formulations and highly complex to solve in a DTA framework, in particular for a realistically large network. Furthermore, none of these formulations include a signal control in the framework while studying the vehicle discharged emission. The emission reduction approaches that consider traffic signal controls are discussed in the following paragraph.

Zhu, Lo, and Lin (2013) presented a post-processing method to estimate emissions at the signalized intersections. A mixed-integer pre-timed traffic signal control was implemented. The emission data for idling mode, acceleration mode, cruising mode, and deceleration mode were coupled with speed to calculate emission factors. A fuel-based signal optimization model was proposed by Liao (2013). This paper presents a model that relates emission with vehicle speed, travel-time, and stop time. It is to mention that the higher the fuel consumption, the greater the emission. Zhou et al. (2015) proposed a simulation framework by incorporating Motor Vehicle Emission Simulator (MOVES) for emission modeling and Newell's simplified kinematic wave model for simulating traffic flow behavior. Then an algorithm was proposed to calculate vehicle trajectory. Traffic signal control was also considered in the proposed simulation method. Chang and Hui (2016) expressed traffic flow regarding impedance functions and formulated a bi-level program where at the upper level the travel-time was minimized and at the lower level, the emission was minimized. The embedded traffic signal control in the framework optimally controls outgoing traffic movements of a phase by controlling saturation flow of the corresponding links. Zhao et al. (2016) proposed a dynamic traffic signal timing algorithm to mitigate travel-time and fuel consumption jointly. A nonlinear fuel consumption model was implemented in the

Table 2.4. The literature related to vehicle discharged emission.

Contributions	**NHB**	**Signal Control**	**Linear**
Benedek and Rilett (1998)	✓	×	×
Sugawara and Niemeier (2002)	✓	×	×
Dirks et al. (2003)	✓	×	×
Zegeye et al. (2009)	✓	×	×
Aziz and Ukkusuri (2012b)	✓	×	×
Zhu, Lo, and Lin (2013)	NA	✓	×
Liao (2013)	✓	✓	×
Friesz et al. (2013)	✓	×	×
Zhu and Ukkusuri (2014)	✓	✓	×
Levin, Duell, and Waller (2014)	✓	×	×
Ma, Ban, and Szeto (2015)	✓	×	×
Liu et al. (2015b)	✓	×	×
Zhou et al. (2015)	NA	✓	×
Wang, Zhang, and Zhang (2016)	✓	×	×
Luo et al. (2016)	✓	×	×
Chang and Hui (2016)	✓	✓	×
Long et al. (2016)	✓	×	×
Zhao et al. (2016)	NA	✓	×
Lu et al. (2016)	✓	×	×
Pasquale et al. (2017)	✓	✓	×
Fernandes, Coelho, and Rouphail (2017)	NA	✓	×
Ma, Ban, and Szeto (2017)	✓	×	×

✓ formulation is/includes; × formulation is not/does not include; Not Applicable (NA).

VISSIM platform. Authors claimed that the fuel consumption can be substantially reduced by implementing the proposed strategic traffic signal control. Fernandes, Coelho, and Rouphail (2017) studied effect of the intersection spacing on emission. In this approach, a simulation platform was introduced by combining VISSIM and vehicle specific power emission methodology. It was found in the paper that closely spaced intersections may result in higher vehicle discharged emission. In all the above papers small networks were studied to validate the models and associated computational efforts were not discussed. Furthermore, all these approaches are based on nonlinear or mixed-integer traffic signal control. Furthermore, all these papers include/propose nonlinear emission factor models.

The existing contributions to reduce vehicle discharged emission discussed in this section are summarized in Table. 2.4. In the table, observe that although there have been several contributions focusing on emission, they are mixed-integer or nonlinear programming based formulations and have very high computational complexity. In this thesis, we propose a unified NHB *Emission-Based Dynamic Traffic Assignment with Signal Control (NHB EB DTA-SC)* formulation. The formulation is linear and jointly minimizes travel-time and emission. This formulation can be used to analyze networks of any size. Details about this combined formulation are outlined in Chapter 3. In the next section, papers that address issues related to public transport traffic assignment are discussed.

2.6 Public Bus Transit With Priority

Improving public transit is one of the most promising strategies to reduce congestion in urban areas. In particular, smarter *Bus Rapid Transit (BRT)* with *Transit Signal Priority (TSP)* can provide the speed and reliability offered by trains/subways as well as flexibility and economy of the buses. Furthermore, it can yield substantially reduced emissions and fuel consumptions. A *Bus Priority System Optimal Dynamic*

Traffic Assignment with Signal Control (BP SO-DTA-SC) mathematical framework would enable us not only to design, plan, strategically manage, and determine policies of the BRT-TSP system, but also to evaluate crucial public transit investments. In the recent past, several contributions report a wide variety of aspects of the public transit network. These approaches either do not consider priority of the transit vehicles or do not include TSP control models. These contributions are reviewed below.

Several studies emphasis on the benefits of BRT. Gan et al. (2003) developed a simulation-based method to justify the design of a dedicated bus lane approach. Viegas and Lu (2004) and Eichler and Daganzo (2006) analyzed the benefits and drawbacks of the intermittent bus lanes which are open to the cars as long as there is no bus in these lanes. Chiu and Zheng (2007) investigated the priority scheme for multiple groups of similar vehicles in the emergency evacuation problem by proposing a linear CTM-based model. Note that their solution also suffers from the unrealistic vehicle HB problem, this issue will be addressed in this research. Li and Ju (2009) presented a point-queue DTA approach to allocate dedicated bus lanes. To the best of our knowledge, this is the only DTA approach which studies dedicated lanes for the public transports. The model is used to reflect interactions between the buses and cars with or without bus exclusive lanes. An integrated Variational Inequality (VI) formulation is proposed to model the discussed problem. Then a heuristic algorithm is developed to solve the VI. A small network was analyzed, but required computational effort was not reported. The approach is based on point queue network loading model. As a result, cannot characterize congestion or spill back scenarios. Furthermore, the presented approach does not include a traffic signal control in the formulation and has very high complexity due to the nonlinear VI formulation. Liu et al. (2015a) proposed an analytical formulation that considers buses as moving bottlenecks due to the slower speed of such vehicles. The conventional CTM is extended for multi-class traffic and named as Bus CTM. The objective function of the proposed formulation minimizes the total system-wide passenger travel time, and to some extent accounts bus priority. Small networks were used to generate numerical solutions. Although computational tractability of the approach was claimed, associated complexity was not reported. Authors showed that with the increasing number of buses the total passenger travel time decreases. Nevertheless, the proposed approach suffers from vehicle Holding-Back problem and does not include a signal control in the formulation. Furthermore, the formulation is mixed-integer and has a dependency on a deterministic parameter which limits its use to the predetermined parameter only. Alonso et al. (2017) proposed a nonlinear traffic assignment model for the bus lanes, and they studied the associated quality of service. However, both these models lack the design of signal control and do not model the give-way behavior (i.e., moving-straight cars have to give-way to the turning buses) in urban networks.

In the literature, there have been several studies focusing on modelling TSP control schemes for the multi-class traffic scenarios without any prior development of DTA models. For example, Wu and Hounsell (1998) introduced an optimal pre-signal control for the buses to have priority access to the downstream junction. Janos and Furth (2002) evaluated priority tactics of signal timing in highly interruptible scenarios in which buses can request to pass through a crowded intersection or to arrive at a bus stop that has long standing queue. Ma, Liu, and Yang (2013) presented an optimal control scheme to minimize the intersection delay for buses without causing excessive delay to other vehicles. Head, Gettman, and Wei (2006) considered not only signal timing but also ring, barrier and phase constructs in the scheme. Ma, Yang, and Liu (2010) studies the adaptive optimal priority strategy based on the estimation of bus delay. Relying on the deployment of communication technology, Chang, Vasudevan, and Su (1996) proposed a performance index that accounts the delay for the passengers and the bus schedules, to adaptively decide on green time extension. Similarly, He, Head, and Ding (2014) proposed a framework for multimodal priority control and resolved the conflict between bus and car movements using an actuated control. Nevertheless, the above contributions did not consider the impact of time-varying traffic where congestion often occurs dynamically and depends naturally on the control decisions. Furthermore, without exception, the existing approaches were all mixed-integer or nonlinear formulations and thus very complex to solve.

Several simulation-based studies have been proposed to address the multi-class traffic scenarios with or without TSP controls. Bliemer (2007) presented a simulation-based DTA model that was focused mainly on the queue development and spillback traffic conditions. He, Head, and Ding (2011) developed a heuristic algorithm to simultaneously process multiple priority requests at an isolated intersection. A microscopic traffic simulator tool was used to analyze the proposed heuristic. Then the proposed approach was compared with other methods of transit vehicle priority. This approach can be applied to the privileged vehicles such as ambulances or transit buses. All these contributions claim noticeable improvement in bus flows at intersections. Nevertheless, simulation-based approaches may attain sub-optimal solutions, and the iterative algorithms impose polynomial solvable complexity.

In summary, none of the presented contributions in the literature studies the combination of all the three components, i.e., *Dedicated Bus-only Lanes with Priority (DBLP)*, TSP, and multi-class traffic behavior in a computationally tractable framework. To fill this gap, we propose a BRT-TSP framework namely BP SO-DTA-SC model. This model is the first DTA based *linear* approach to formulate a *BRT-TSP mathematical framework* and can be applied to analyze large scale transit priority systems. Moreover, the proposed formulation considers bus routes, obtains NHB solutions, and assures fairness to all the passengers so that bus priority does not impose excessive delay to the car users.

2.7 Chapter Summary

At the beginning of this chapter, a literature review related to SO-DTA with/without signal control model was outlined. Then the relevant existing contributions associated with the properties of the optimal solutions of the DTA frameworks were discussed. Existing traffic assignment and signal control approaches that reduce vehicle discharged emissions were also explored. Finally, papers related to BRT-TSP models were reviewed. The key research gaps that were found in the literature review are summarized below.

- A computationally tractable linear NHB SO DTA-SC formulation is yet to be developed.
- Structural properties of the optimal traffic assignment and signal control policy in an NHB SO setting is not yet explored.
- A unified traffic assignment and signal control framework that jointly minimizes emission and travel-time is yet to be introduced.
- A computationally tractable BRT-TSP mathematical model is not yet presented in the literature.

To fill these research gaps, we propose and analyze

- an NHB DTA-SC formulation in Chapter 3,
- a linear EB DTA-SC formulation in Chapter 3,
- several structures of the optimal solution in Chapter 4, and
- a novel linear and NHB BP SO-DTA-SC formulation in Chapter 5.

Chapter 3

Non-Holding-Back Dynamic Traffic Assignment Formulations

3.1 Introduction

It is evident that the majority of the Non-Holding-Back (NHB) DTA frameworks with system optimal objective (SO-DTA) presented in the literature are based on mixed-integer (MI) or non-linear formulations with high complexity and computational burden. Similarly, most of the existing signal control models that can be embedded in a DTA framework are mixed-integer or non-linear formulations or trade accuracy for a realistic Cycle-Length (CL). Furthermore, all the emission based DTA formulations in the literature are non-linear and have very high computational complexity. To address these issues, we focus on developing realistic and computationally tractable formulations. The main contributions of this chapter are as follows:

- We propose a linear and continuous NHB DTA-SC formulation for the single destination networks which attains SO solution. We prove that this formulation is completely free of the unrealistic vehicle holding back problem.
- The embedded linear signal control model in the NHB DTA-SC formulation allows setting practically long CL (e.g., 60s) for any short discrete time-slot (e.g., 5s). We name it as Signal Control with Realistic Cycle-length (SCRC). The shorter time-scale is required to improve the accuracy of the solution and

its usability for road segments of any length.

- We numerically show that the time-scale of a DTA problem has a significant impact on traffic dynamics (e.g., density, travel time) and that a shorter time-scale more accurately locates the position of bottlenecks or congestion on the links.

- We further investigate the time-scale issue from the emission perspective. To estimate emission, we propose two approaches, that are,

  – A method to estimate emission from the obtained DTA solution by combining vehicle-speed and an emission model.

  – A linear Emission-Based DTA with Signal Control (EB DTA-SC) formulation. This formulation is HB free and assigns traffic to attain the lowest possible emission for deterministic demands. Though the EB DTA-SC is formulated to minimize NO_X, it can be easily extended to other pollutants (e.g., CO).

The remainder of the chapter is organized as follows. In Section 3.2 the holding back problem in the SO-DTA formulation is highlighted. Then a solution to the Holding-Back problem is proposed. A novel linear-continuous traffic signal control with realistic cycle time is presented in this section followed by the emission estimation methods. Section 3.3 presents results and analysis of the solutions, and Section 3.3 concludes this chapter by summarizing the key outcomes of this study.

3.2 Model Formulation

In a DTA framework, the Dynamic Network Loading (DNL) model specifies how traffic propagates over a given network in space and time. In this research, the widely accepted Cell Transmission Model (CTM) proposed by Daganzo (1994, 1995) is used as the DNL model. The CTM is the discrete analogue of the hydrodynamic flow-density differential equations. Compared to other second order macroscopic DNL models such as METANET (Messmer and Papageorgiou 2000), CTM provides relatively realistic details about queue formation, propagation, and dissipation of congestion through kinematic waves and is simpler for urban model based traffic control purposes (Nie 2006). In CTM, the road segments are divided into a number of homogeneous units called cells. The cell length and the discrete time slot are intrinsically related to each other and are chosen to satisfy the Courant-Friedrichs-Lewy (CFL) condition (LeVeque 1992). It is known that the accuracy of CTM increases as the discrete time slots become shorter (Munoz et al. 2003; Sun and Bayen 2008).

3.2.1 The HB Problem in the Linearized CTM

In the conventional SO-DTA formulation with CTM (Ziliaskopoulos 2000), the vehicle Holding-Back (HB) problem arises due to the relaxation of the non-linear CTM equations to a set of linear constraints. Though this HB SO-DTA formulation is tractable for realistic size networks, the HB problem limits its application. The list of notations used throughout this dissertation is given below. The HB SO-DTA formulation (P_Z) presented by Ziliaskopoulos (2000) is as follows:

Sets

$\mathcal{T}$	set of all discrete time intervals t
$\mathcal{T}_c$	set of all the signal control cycle index c
$\mathcal{C}$	set of all the cells
$\mathcal{C}_D$	set of all the diverging cells
$\mathcal{C}_M$	set of all the merge cells
$\mathcal{C}_{IM}$	set of all the intersection merge cells
$\mathcal{C}_O$	set of all the ordinary cells
$\mathcal{C}_S$	set of all the sink cells
$\mathcal{C}_R$	set of all the source cells
$\mathcal{E}$	set of all the cell connectors
$\mathcal{E}_R$	set of $\mathcal{C}_R$ connectors
$\mathcal{E}_O$	set of $\mathcal{C}_O$ connectors
$\mathcal{E}_M$	set of $\mathcal{C}_M$ connectors

$\mathcal{E}_D$	set of $\mathcal{C}_D$ connectors
$\mathcal{C}_I$	set of all the intersection cells
$\mathcal{E}_S$	set of all the sink connectors
$\mathcal{E}_{IM}$	set of $\mathcal{C}_{IM}$ connectors
$\Gamma^{-1}(i)$	set of all the predecessor cells of cell i
$\Gamma(i)$	set of all the successor cells of cell i
$\mathcal{I}$	set of all the intersections where I is an intersection $\in \mathcal{I}$
P	set of all the signal phases
Φ	set of all the traffic movements at an intersection

Index, parameters, and variables

τ	duration of each time-slot
t	a time-slot
p	phase within a traffic cycle
ϕ_h	a traffic movement at an intersection
c	traffic signal control cycle index
i, j, k	cell index
q	flow
ρ	density
ρ_{cd}	critical density
ρ_{jd}	jam density
V_i^t	free-flow speed at cell i at time-slot t
v_i^t	average-speed of the vehicles at cell i at time-slot t
w_i^t	backward speed at cell i at time-slot t
δ_i^t	the ratio of free-flow speed and backward propagation speed V_i^t/w_i^t
d_i^t	demand at the source cell i at time-slot t
N_i^t	occupancy capacity of cell i at time-slot t
Q_i^t	maximum flow into/out of cell i at time-slot t
g_{min}	minimum allowable green time
L	length of a cell
$EFNO_{X_i}^t$	nitrogen oxide (NOx) emission factor at cell i and time-slot t
E_{NO_X}	total system-wide emission (TSE)
Ψ_i	artificial penalty on occupancy of cell i
Θ_i	artificial incentive on occupancy of cell i
x_i^t	occupancy of cell i at time-slot t
y_{ij}^t	flow from cell i to cell j at time-slot t
g_i^{c,ϕ_p}	green-split of cell i assigned for the traffic movement ϕ_p in phase p of a cycle c
κ_{ij}^t	1, if the movement is allowed from cell i to j at slot t
ξ_p^t	1, if the phase p is active at slot t

P_Z :

$$\text{minimize} \quad \mathcal{O}_Z = \sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C} \setminus \mathcal{C}_S} x_i^t \tag{3.1}$$

subject to

$$x_i^t = x_i^{t-1} + d_i^{t-1} - y_{ij}^{t-1}, \ \forall i \in \mathcal{C}_R, \ j \in \Gamma(i), \ \forall t \in \mathcal{T} \tag{3.2}$$

$$x_i^t = x_i^{t-1} + y_{ki}^{t-1} - y_{ij}^{t-1}, \ \forall (i,j) \in \mathcal{E}_O, \ j \in \Gamma(i), \ \forall t \in \mathcal{T} \tag{3.3}$$

$$x_i^t = x_i^{t-1} + y_{ki}^{t-1} - \sum_{j \in \Gamma(i)} y_{ij}^{t-1}, \ \forall i \in \mathcal{C}_D, \ k \in \Gamma^{-1}(i), \ \forall t \in \mathcal{T} \tag{3.4}$$

$$x_i^t = x_i^{t-1} + \sum_{k \in \Gamma^{-1}(i)} y_{ki}^{t-1} - y_{ij}^{t-1}, \ \forall i \in \mathcal{C}_M \cup \mathcal{C}_{IM}, \ \forall t \in \mathcal{T} \tag{3.5}$$

$$y_{ij}^t - x_i^t \leq 0, \ y_{ij}^t \leq Q_i^t, \ \forall (i,j) \in \mathcal{E}_{R \cup O}, \ j \in \Gamma(i), \ \forall t \in \mathcal{T} \tag{3.6}$$

$$y_{ij}^t \leq Q_j^t, \ y_{ij}^t \leq \delta_j^t (N_j^t - x_j^t), \ \forall (i,j) \in \mathcal{E}_{R \cup O \cup D}, \ j \in \Gamma(i), \ \forall t \in \mathcal{T} \tag{3.7}$$

$$\sum_{j \in \Gamma(i)} y_{ij}^t - x_i^t \leq 0, \ \sum_{j \in \Gamma(i)} y_{ij}^t \leq Q_i^t, \ \forall (i,j) \in \mathcal{E}_D, \ j \in \Gamma(i), \ \forall t \in \mathcal{T} \tag{3.8}$$

$$y_{ki}^t \leq Q_k^t, \ \forall (k,i) \in \mathcal{E}_{M \cup S}, \ k \in \Gamma^{-1}(i), \ \forall t \in \mathcal{T} \tag{3.9}$$

$$y_{ki}^t \leq x_k^t, \ \forall (k,i) \in \mathcal{E}_{M \cup S \cup IM}, \ k \in \Gamma^{-1}(i), \ \forall t \in \mathcal{T} \tag{3.10}$$

$$\sum_{k \in \Gamma^{-1}(i)} y_{ki}^t \leq \delta_i^t (N_i^t - x_i^t), \ \forall (k,i) \in \mathcal{E}_M \cup \mathcal{E}_{IM}, \ k \in \Gamma^{-1}(i) \tag{3.11}$$

$$\sum_{k \in \Gamma^{-1}(i)} y_{ki}^t \leq Q_i^t, \ \forall (k,i) \in \mathcal{E}_M \cup \mathcal{E}_{IM}, \ k \in \Gamma^{-1}(i), \ \forall t \in \mathcal{T}. \tag{3.12}$$

The objective function (3.1) minimizes occupancy of the vehicles throughout the whole network, which is equivalent to minimizing Total System-wide Travel Time (TSTT) (Ziliaskopoulos 2000). Equations (3.2)-(3.5) represent the linear conservation equations or the flow-occupancy balance relations for different types of cells, and constraints (3.6)-(3.12) are linear relaxations to the non-linear CTM flow propagation rules introduced by Daganzo (1994). For example, the non-linear CTM flow propagation rule for ordinary cells is given by

$$y_{ij}^t = \min \left\{ x_i^t, \min[Q_i^t, Q_j^t], \delta_j^t (N_j^t - x_j^t) \right\}, \ \forall (i,j) \in \mathcal{E}_O, \ j \in \Gamma(i), \ \forall t \in \mathcal{T}. \tag{3.13}$$

Note that the non-linearity in Equation (3.13) is due to the 'min' operator. The vehicle Holding-Back problem arises precisely due to the relaxation of the non-linear CTM flow propagation rules; if for a flow y_{ij}^t, none of the inequalities in constraints (3.6)-(3.7) holds with equality, then Equation (3.13) will not hold with equality, which results in Holding-Back of vehicles at cell i. The traffic propagation over the links pertaining to an SO solution with HB does not represent real-life traffic dynamics. Further details about the Holding-Back problem associated with the above formulation can be found in Zhu and Ukkusuri (2013) and Doan and Ukkusuri (2012). For detailed explanations of Equations (3.2)-(3.13), see Ziliaskopoulos (2000).

3.2.2 The Proposed NHB SO-DTA Framework

We propose a linear objective function that can successfully resolve the vehicle Holding-Back (HB) problem. Our proposed objective is given by

$$\mathcal{O} = \max \sum_{\forall t \in \mathcal{T}} \sum_{\forall ij \in \mathcal{E}_S} y_{ij}^t - \sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C} \setminus \mathcal{C}_S} \Psi_i \cdot x_i^t. \tag{3.14}$$

The first term in the above objective represents the total flow into the sink cells, hence, its maximization is equivalent to the maximization of the total network-wide throughput. The second term represents a penalty where an artificial penalty Ψ_i is multiplied with the occupancy at each cell i. The idea is that the HB problem can be eliminated by adding a cell-dependent penalty at each cell. The penalty Ψ_i is predefined. The highest penalty is associated with the source cells, and the penalty decreases as one moves closer towards the destination. If cell i is upstream to cell j and they are on the same link, then $\Psi_i > \Psi_j$. This is to mention that the second term minimizes occupancy throughout the whole network (i.e., travel-time) while the penalty terms work as incentives to the objective function to move vehicles towards the destination regardless of the size of the cell or the network. This property of the objective function necessarily eliminates the HB problem. Further elaboration is provided in Proposition 3.2.3. The penalty terms can be generated by a simple algorithm outlined in Section 3.2.2.1. A similar algorithm was presented by Zhu and

Ukkusuri (2013).

Remark 1. *Note that the algorithm presented by Zhu and Ukkusuri (2013) applies penalty terms from the destination cell to the source cell. In their approach paths are assumed to be defined from source to destination, and penalty terms are computed based on the relative positions of the cells from the destination cell. As a result, in a cyclic network use-case, this approach applies higher penalties for the cells in the loop compared to the non-loop cells although the formers are closer to the destination cell. In a penalty based NHB framework, this might result in suboptimal solutions. To overcome this limitation, we explicitly define the paths and assign penalty terms based on these paths rather than cells. Our proposed algorithm can be applied to determine penalty terms for symmetric and cyclic networks. An example of the mentioned case is presented in the Appendix A.1.* $\odot$

3.2.2.1 Algorithm to Determine Penalty Terms (Ψ_is)

Notes:

- This algorithm is designed for *single destination CTM networks.* In the CTM, a connector that connects a diverging and a merging cell is not allowed (Daganzo 1995; Ziliaskopoulos 2000).
- Here i and j are cell indexes, and $i \in \Gamma^{-1}(j)$.
- $\mathcal{P}$ is the set of all the paths.

Definition 3.2.1. (Definition of path for determining Ψ_i values) A path $p_{a(S \to R)}$ is an ordered list of cells starting from the sink cell and goes towards a source cell by facing each of the cell connector (directional) from the opposite direction where $a(S \to R)$ is the path-index. If a path

- $p_{a(S \to R)}$ consists of *no cell more than once*, then the path ends at a source cell R;

– $p_{a(S\to R)}$ *repeats a cell* j *for the second time* while going towards the source cell, then the path terminates at cell i.

$\triangle$

Start

1. Initialize:

   (a) $\Psi_i = 1, i \in \mathcal{C}_S$.

   (b) Determine set of all the paths $\mathcal{P}$ in which each path follows Definition 3.2.1.

2. Determine path-based penalty terms using following relation. Start from predecessor-cell of the sink cell and move towards the source cell. Stop when a path ends (i.e. the last cell in the path). Repeat for other paths.

   - For $\forall p_{a(S\to R)} \in \mathcal{P}$: $\Psi_{ip_{a(S\to R)}} = \Psi_{jp_{a(S\to R)}} + 1$, when, $i \notin \mathcal{C}_S$.

3. Convert all the path specific penalty terms to cell specific penalty terms using following relation, i.e., if a cell has two or more path specific penalty terms, compare them and select the maximum value; if a cell has just one path specific penalty term then take it as the cell specific penalty.

   - $\Psi_i = \max(\Psi_{ip_{a(S\to R)}}),\ i \notin \mathcal{C}_S,\ \forall p_{a(S\to R)} \in \mathcal{P}$.

End

3.2.2.2 System Optimal (SO) Property

To ensure the SO property, we add the following constraint

$$\sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C} \setminus \mathcal{C}_S} x_i^t = \mathcal{O}_Z^\star, \tag{3.15}$$

where $\mathcal{O}_Z^\star$ is the objective value pertaining to the HB SO-DTA formulation proposed by Ziliaskopoulos (2000) for the same network. Thus, our proposed NHB formulation is obtained from the SO-DTA formulation in P_Z by replacing Equation (3.1) with (3.14) and adding (3.15) as an additional constraint.

Zhu and Ukkusuri (2013) present a similar linear NHB objective. However, the formulation in Zhu and Ukkusuri (2013) requires one to choose certain parameters M_L and M_S, and may not attain a solution for a wrong choice of these parameters. On the other hand, to be able to guarantee a SO solution, one needs to choose specific values of the parameters M_L and M_S that requires the knowledge of the optimal solution of the problem itself, thereby leading to a cyclic dependency. Our proposed approach overcomes these limitations.

3.2.2.3 Proof of SO and NHB Property of the Solution

It is proven by Shen, Nie, and Zhang (2007) that a Holding-Back SO solution always has a Non-Holding-Back SO solution for the same optimal objective value. Since the equality constraint (3.15) is a part of the proposed NHB SO-DTA formulation, it guarantees that the solution is indeed SO. We formally write it as a proposition.

Proposition 3.2.2. *If $\bar{x}$ is an optimal solution of the proposed NHB SO-DTA framework, then this solution is System Optimal (SO), i.e., for all other feasible x, we have $\sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C} \setminus \mathcal{C}_S} \bar{x}_i^t \leq \sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C} \setminus \mathcal{C}_S} x_i^t$.*

Next, we show that the proposed NHB objective in (3.14) can indeed resolve the Holding-Back problem.

Proposition 3.2.3. *The optimal solution of the proposed NHB objective does not have any holding back if the optimization is solved with a time horizon $\mathcal{T}$ that is long enough for all the traffic demand to be cleared.*

Proof. Suppose that the optimal solution $(\bar{x}, \bar{y})$ with the NHB objective holds back traffic somewhere within the network. A cell connector (i, j) belonging to some ordinary, diverge, or merge cell is involved in the Holding-Back (HB). Assume that the Holding-Back occurs at cell i and let Δ denote the amount of HB traffic in cell i at time slot $t+1$. Then, one of the following strict inequalities, $\mathcal{E}_O, \mathcal{E}_M$ or $\mathcal{E}_D$, must hold at the time slot t.

$$\exists(i,j)\ \in \mathcal{E}_O,\ \bar{y}_{ij}^t < \min\left\{\bar{x}_i^t,\ Q_i^t,\ Q_j^t, \delta_j^t\left(N_j^t - \bar{x}_j^t\right)\right\} \tag{3.16}$$

or

$$\exists(i,j)\ \in \mathcal{E}_M,\ \bar{y}_{ij}^t < \min\left\{\bar{x}_i^t, Q_i^t\right\} \text{ and } \sum_{m\in\Gamma_j^{-1}} \bar{y}_{mj}^t < \min\left\{Q_j^t, \delta_j^t\left(N_j^t - \bar{x}_j^t\right)\right\} \tag{3.17}$$

or

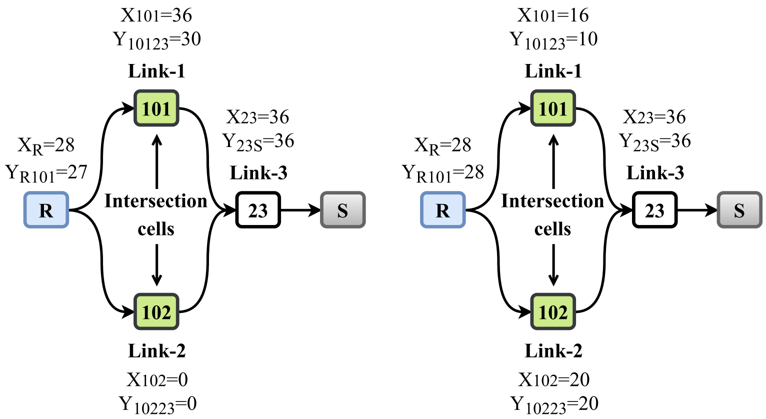
$$\exists(i,j)\ \in \mathcal{E}_D,\ \sum_{m\in\Gamma_i} \bar{y}_{im}^t < \min\left\{\bar{x}_i^t,\ Q_i^t\right\} \text{ and } \bar{y}_{ij}^t < \min\left\{\ Q_j^t, \delta_j^t\left(N_j^t - \bar{x}_j^t\right)\right\}. \tag{3.18}$$

For the sake of simplicity let us assume that the Holding-Back occurs at an ordinary cell. Therefore, it is possible to move Δ amount of traffic from the ordinary cell i to some downstream cell j, at time t and keep all other cell occupancies unchanged. Let $(x^\star,\ y^\star)$ be the NHB solution obtained by moving this Holding-Back traffic. Let $\bar{\mathcal{O}}$ and $\mathcal{O}^\star$ denote the objective values for the solutions $(\bar{x}, \bar{y})$ and $(x^\star, y^\star)$, respectively. Then, one can write

$$\begin{aligned}\mathcal{O}^\star - \bar{\mathcal{O}} = &\sum_{\forall t\in\mathcal{T}} \sum_{\forall ij\in\mathcal{E}_S} y_{ij}^{\star t} - \sum_{\forall t\in\mathcal{T}} \sum_{\forall i\in\mathcal{C}\setminus\mathcal{C}_S} \Psi_i \cdot x_i^{\star t} \\ &- \sum_{\forall t\in\mathcal{T}} \sum_{\forall ij\in\mathcal{E}_S} \bar{y}_{ij}^t + \sum_{\forall t\in\mathcal{T}} \sum_{\forall i\in\mathcal{C}\setminus\mathcal{C}_S} \Psi_i \cdot \bar{x}_i^t\end{aligned} \tag{3.19}$$

Let the total demand generated at the source cell be d. Since all the traffic demands reach the destination within $\mathcal{T}$, it follows that

$$\sum_{\forall t\in\mathcal{T}} \sum_{\forall ij\in\mathcal{E}_S} y_{ij}^{\star t} = \sum_{\forall t\in\mathcal{T}} \sum_{\forall ij\in\mathcal{E}_S} \bar{y}_{ij}^t = d \tag{3.20}$$



(a) The HB SO-DTA solution. (b) The NHB SO-DTA solution.

Figure 3.1. Figure shows optimal occupancies ($\mathrm{X_i}$) and flows ($\mathrm{Y_{ij}}$) attained using the HB and NHB formulation at the time slot $t = 4$ where i is the cell index and j is the index of the successor cell. In the HB SO-DTA case, one vehicle is being held back at the source cell R. Whereas, this problem is resolved in the NHB SO-DTA case.

and Equality 3.19 simplifies to

$$\begin{aligned}\mathcal{O}^{\star} - \bar{\mathcal{O}} &= \sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C} \setminus \mathcal{C}_S} \Psi_i \cdot \bar{x}_i^t - \sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C} \setminus \mathcal{C}_S} \Psi_i \cdot x_i^{\star t} \\ &= \Psi_i \bar{x}_i^t + \Psi_j \bar{x}_j^{t+1} - \Psi_i x_i^{\star t} - \Psi_j x_j^{\star t+1} \\ &= \Psi_i \left(\bar{x}_i^t - x_i^{\star t} \right) - \Psi_j \left(x_j^{\star t+1} - \bar{x}_j^{t+1} \right) \\ &= \left(\Psi_i - \Psi_j \right) \Delta \end{aligned} \tag{3.21}$$

If $\Delta > 0$, then we have $\mathcal{O}^{\star} - \bar{\mathcal{O}} > 0$, because $\Psi_i > \Psi_j$. This will be a contradiction since $\bar{\mathcal{O}}$ is, by definition, the solution of the maximization objective (3.14). Therefore, $\Delta = 0$ must hold, which implies that there is no holding back.

The proof for the general case is a straightforward extension and involves moving the Δ amount of traffic between multiple cells. □

To illustrate the proposed NHB formulation discussed above, we consider the 60s CTM network in Figure 5(a) with the physical parameters in Table. 3.2. Since, our purpose here is to show that the proposed objective can resolve vehicle Holding-Back problem, we apply the demand of $d_R^1 = 100$ only at one time-slot $t = 1$. In Figure 3.1, the optimal cell occupancies and flows are shown for both the Holding-Back SO-DTA and NHB SO-DTA solutions. In Figure 3.1 observe that the HB solution holds 1 vehicle at $t = 4$ at the source cell R and sends 27 vehicles to the cell 101. At this time, cell 101 has available occupancy capacity for another 36 vehicles while cell 102 is empty. Therefore, the vehicle should not have been held back at the source cell. On the other hand, in the NHB solution the mentioned vehicle indeed moves to the successor cell.

3.2.3 Signal Control Models

In this section, the constraints related to the signal control models are presented. Along with the proposed Signal Control with Realistic Cycle-length (SCRC) model, the cycle-length Same as Discrete Time-interval (CSDT) model (Ukkusuri, Ramadurai, and Patil 2010) and the Mixed Integer Signal Control (MISC) model (Aziz and Ukkusuri 2011) are also discussed for comparison purposes.

3.2.3.1 The Proposed SCRC Model

The underlying idea of the SCRC model is that the optimal green splits are computed as a fraction of the signal control Cycle-Length (CL). Each signal control cycle-length consists of one or more CTM time-slots. As a result, the proposed SCRC model allows us to set the signal control CL to any multiple time-slots and yet retains the linearity of the SO-DTA formulation.

Throughout this thesis the cycle-index is denoted by c. To derive the relationship between the signal control cycle-index c and the CTM time-slot index t, let us assume that the signal control cycle consists of m discrete time-slots. The cycle-index c can be written as

$$c = \lfloor \frac{t-1}{m} + 1 \rfloor, \tag{3.22}$$

where, c and m denote cycle-index and cycle-length, respectively. An example is provided in Figure 3.2 with the cycle-length of $m = 3$ time-slots. The time-slot index

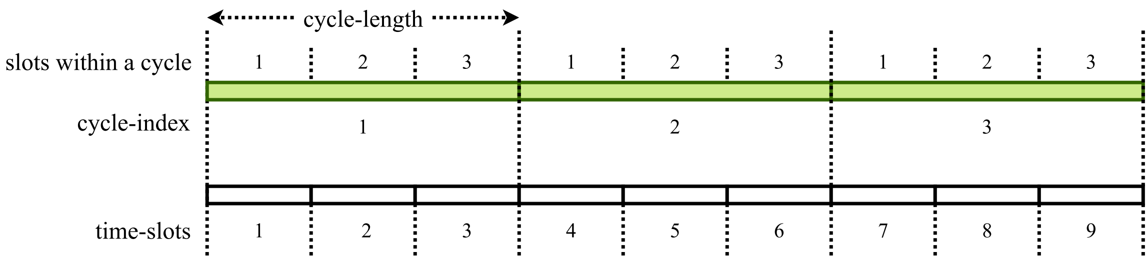


Figure 3.2. The relation between SCRC cycle-index and time-slot index when the cycle-length $m = 3$ time-slots, which is three time-slots are combined to form the cycle-length. The slots within a cycle is also visible in the figure. The green-split optimization decision variable g_i^{c,ϕ_p} is computed over the whole cycle-length of m time slots.

t within the cycle-index c can be determined by

$$t = (c-1)m + \varepsilon_c,\ \forall \varepsilon_c \in \{1, ..., m\},\ \forall c \in \mathcal{T}_c,\ \forall t \in \mathcal{T}, \tag{3.23}$$

where, ε_c denotes a slot index within the c signal control cycle.

In each cycle, a traffic phase p ($p \in P$) allows several non-conflicting traffic movements to pass through the intersection. Herein, we use the signal phasing convention prescribed by the national electrical manufacturers association (NEMA) which has four phases denoted in this research as $P = \{1, 2, 3, 4\}$. Let $\Phi = \{\phi_h\},\ h \in \{1, 2, 3, ..., 12\}$ be the set of all the possible outgoing traffic movements from the intersection cells. The NEMA convention mandates that the (ϕ_p, ϕ_{p+4}) and (ϕ_{2p}, ϕ_{p+8}) traffic movement pairs are allowed together where $p \in P$ represents a phase in a cycle.

Let g_i^{c,ϕ_p} be the green split allocated to the phase $p \in P$ which enables the traffic movement ϕ_p from an intersection cell i, $i \in \mathcal{C}_I$ in the cycle c. Note that because of the allowable traffic movement pairs discussed above, ϕ_p covers all the possible outgoing traffic movements at the intersection.

The green split g_i^{c,ϕ_p} restricts the maximum flow capacity Q_i^t for the traffic from the intersection cell i, $i \in \mathcal{C}_I$ over the CL period, and leads to the constraint

$$y_{ij}^t \leq g_i^{c,\phi_p} Q_i^t,\ \forall i \in \mathcal{C}_I, j \in \Gamma(i), c \in \mathcal{T}_c, \phi_p \in \Phi, t \in \mathcal{T}. \tag{3.24}$$

The NEMA convention leads to the following constraints

$$g_i^{c,\phi_p} = g_i^{c,\phi_{p+4}},\ \forall c \in \mathcal{T}_c,\ p \in P,\ i \in \mathcal{C}_I. \tag{3.25}$$

$$g_i^{c,\phi_{2p}} = g_i^{c,\phi_{p+8}},\ \forall c \in \mathcal{T}_c,\ p \in P,\ i \in \mathcal{C}_I. \tag{3.26}$$

The total green time for any phase should not be below a minimum value, which leads to the constraint

$$g_i^{c,\phi_p} \geqslant g_{min},\ \forall c \in \mathcal{T}_c,\ p \in P,\ i \in \mathcal{C}_I. \tag{3.27}$$

Finally, one must make sure that the sum of the green fractions cannot exceed 1, which can be mathematically written as

$$\sum_{p \in P} g_i^{c,\phi_p} \leqslant 1,\ \forall c \in \mathcal{T}_c,\ p \in P,\ i \in \mathcal{C}_I. \tag{3.28}$$

3.2.3.2 The CSDT Signal Control Model

The idea of the CSDT model was first discussed by Daganzo (1994) and later integrated into a linear continuous DTA formulation by Ukkusuri, Ramadurai, and Patil (2010). In this model, a green split variable ω_i^t is maintained for each intersection cell i and for each discrete time slot t. The length of the discrete time slot must be equal to the cycle time. Therefore, for realistic cycle times of the order of 60-120 sec, the cell length in the CTM model is of the order of 1-2 km. Such long cells may not always be possible to create, especially for short urban links, and may compromise the accuracy and usability of the solution.

The effect of the variable ω_i^t in the CSDT model is similar to that of g_i^{c,ϕ_p} in the proposed SCRC model, and leads to the constraint

$$y_{ij}^t \leq \omega_i^t Q_i^t,\ i \in \mathcal{C}_I,\ j \in \Gamma(i),\ \forall (i,j) \in \mathcal{E}_{IM},\ \forall t \in \mathcal{T}. \tag{3.29}$$

Detailed formulation of the model including the definition of the phases can be found in Ukkusuri, Ramadurai, and Patil (2010).

3.2.3.3 The MISC Signal Control Model

The MISC model was presented by Aziz and Ukkusuri (2011). In this model, an indicator variable κ_{ij}^t is maintained for each movement ij where $\kappa_{ij}^t = 1$ if the traffic movement ij is allowed at time t, and $\kappa_{ij}^t = 0$ otherwise. The effect of the variable κ_{ij}^t in the MISC model is similar to that of g_i^{c,ϕ_p} in the proposed SCRC model and leads to the constraint

$$y_{ij}^t \leq \kappa_{ij}^t Q_i^t, \quad i \in \mathcal{C}_I,\ j \in \Gamma(i),\ \forall (i,j) \in \mathcal{E}_{IM},\ \forall t \in \mathcal{T}. \tag{3.30}$$

At an intersection only one phase should be active at any time interval. This is assured by setting

$$\sum_{p \in P} \xi_p^t = 1 \ \ , \ \ \forall t \in \mathcal{T}, \tag{3.31}$$

where ξ_p^t is another 0-1 indicator variable and activates one phase at a time. To integrate flow propagation with the active phase, the following relation between ξ_p^t and κ_{ij}^t is used.

$$\kappa_{ij}^t = \sum_{p \in \sigma_{ij}} \xi_p^t, \quad \sigma_{ij} \subset P, \quad \forall t \in \mathcal{T}. \tag{3.32}$$

In (3.32), σ_{ij} is the set of phases in which the movement ij is activated. For defining the cycle-length, the maximum green time per cycle is kept below or equal to m number of slots per cycle, which leads to the constraint

$$\sum_{\widetilde{t}=t}^{t-m} \xi_p^{\widetilde{t}} \leq m, \quad \forall t \in \mathcal{T}. \tag{3.33}$$

3.2.4 The Proposed NHB DTA-SC Framework

Our complete NHB SO-DTA formulation with Signal Control (SC) (NHB DTA-SC) is given below.

$P1$:

$$\text{maximize} \quad \mathcal{O}_{NHB} = \sum_{\forall t \in \mathcal{T}} \sum_{\forall ij \in \mathcal{E}_S} y_{ij}^t - \sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C} \setminus \mathcal{C}_S} \Psi_i \cdot x_i^t$$

subject to Equations (3.2)-(3.12), (3.24), (3.25)-(3.28), and

$$\sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C} \setminus \mathcal{C}_S} x_i^t = \hat{\mathcal{O}}_Z, \tag{3.34}$$

where $\hat{\mathcal{O}}_Z$ is attained using Equations (3.1)-(3.12), (3.24), (3.25)-(3.28),

and the non-negativity constraints $y_{ij}^t \geq 0,\ x_i^t \geq 0,\ g_i^{c,\phi_p} \geq 0$. In this form, the NHB DTA-SC formulation can be applied to analyze single destination networks such as car parkings or freeway ramps. The above formulation can be extended for the multiple origin-destination networks by amending the models to define paths. In this case, cell-specific penalty terms should be determined for every path going through a cell and the cell would have multiple penalty terms for each of the paths. For the comparison of the results, we have also implemented the CSDT and the MISC models discussed in the previous section.

3.2.5 Emission Estimation Using an Optimal Solution

The traffic assignment in the NHB SO-DTA framework closely follows real-life traffic evolution characteristics (e.g., speed, density). Such traffic propagation is more realistic than that of the Holding-Back (HB) SO-DTA case. Furthermore, the NHB SO-DTA attains entirely different solution than the HB SO-DTA. It is to mention that the optimal flow and density reached using both of the formulations change with shorter or larger CTM time-slots. Since traffic density and speed determines the amount of vehicle induced emission, we are also interested in studying the impact of time-scale on both of the discussed formulations and emission. The key element to estimate emission is speed. This is because the emission factor (EF) is directly related to the speed of the vehicles.

A seminal speed-density relationship was proposed by Greenshields et al. (1935). This model was developed using empirical data. Since then this model has been widely used to derive the fundamental diagram of road traffic such as the approach presented by Pipes (1967). Due to the macroscopic nature of the Greenshields et al. (1935)

speed-density model, it can be easily applied to the CTM use-case. Furthermore, the model is linear and computationally tractable. The model is as follows,

$$v = V(1 - \frac{\rho}{\rho_{jd}}). \tag{3.35}$$

In the above equation v is the average speed of the vehicles in kilometers per hour (km/hr), V is the free-flow speed (km/hr), ρ_{jd} is the jam-density in vehicles per kilometer (veh/km), and ρ is the average density of the vehicles in veh/km.

From the above formula, we derive the average speed of the vehicles in a cell i in terms of CTM parameters and optimal occupancy. Let τ denote the length of the CTM time-slot in seconds (s). Both the free-flow speed V and the average-speed v_i^t in meters per slot (m/slot). The homogeneous cell-length L in meters (i.e. $\tau \cdot V$), and both the jam-density N_i^t and the average-density x_i^t of the vehicles in a cell at a time-slot t in vehicles per cell-length (veh/cell-length). For the CTM use-case, we amend the Equation (3.35) as follows,

$$v_i^t = \frac{V}{\tau}(1 - \frac{\frac{x_i^t}{L}}{\frac{N_i^t}{L}}) \cdot \frac{3600}{1000} = \frac{V}{\tau}(1 - \frac{x_i^t}{N_i^t}) \cdot \frac{3600}{1000} \quad \text{(km/hr)}, \ \forall i \in \mathcal{C} \setminus \mathcal{C}_S, \ \forall t \in \mathcal{T}. \tag{3.36}$$

In the above formula, V is the free-flow speed in any cell of the network when homogeneous cell length is considered throughout the network. x_i^t and v_i^t are the optimal occupancy and their corresponding average-speed in cell i at time-slot t, respectively. All other variables and parameters are same as that of the NHB SO-DTA formulation.

To compute the nitrogen oxide (NO_X) emission factor (EF), we use the empirical model presented by Xia and Shao (2005). This model was derived from empirical data for the small cars. We adapt this model to make it applicable to a DTA formulation. The amended model is presented below,

$$EFNO_{X_i}^t = 0.000247(v_i^t)^2 + 0.0014 v_i^t + 1.387 \quad \text{(gm/km)}, \ \forall i \in \mathcal{C} \setminus \mathcal{C}_S, \ \forall t \in \mathcal{T}. \tag{3.37}$$

In the above formula, the $EFNO_{X_i}^t$ is the NO_X emission factor at cell i at time-slot t and v_i^t is the average-speed of the vehicles at that time-slot in that cell. In the equation, the unit of the EF factor is gm/km per vehicle and that of the speed is km/hr.

The EF is calculated by combining Equation (3.37) and (3.36) for the each of the cell i at the each of the time-slot t in a network. Then, the EFs are substituted into the following equation to obtain the Total System-wide Emission (TSE) E_{NO_X}.

$$E_{NO_X} = \sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C} \setminus \mathcal{C}_S} x_i^t \cdot \frac{EFNO_{X_i}^t \cdot L}{1000} \quad \text{(gm)}. \tag{3.38}$$

In Equation (3.38), L is the cell-length (m) and x_i^t is the optimal occupancy in the NHB SO-DTA or NHB DTA-SC solution. Please notice in the equation that the emission factor is divided by 1000. This is to convert the EF to gm/m per vehicle. As a result, the unit of TSE becomes grams (gm).

3.2.6 A Linear EB DTA-SC Formulation

In the previous section, we have shown how one can estimate emission at the post processing stage by using an optimal solution of a DTA formulation. In this section, we propose an Emission-Based Dynamic Traffic Assignment with Signal Control (EB DTA-SC) formulation that assigns traffic to attain lowest possible emission. The fundamental idea of this formulation is that if one can minimize emission factor at the each of the cells at every time-slot, then the total network-wide emission is also minimized. The proposed EB DTA-SC formulation is linear and computationally tractable. To formulate an objective with emission consideration, let us recall the $EFNO_X$ model in Equation (3.37). This model is non-linear. As a result, it is computationally very expensive. Attaining optimal solution using this model for a reasonable size network would require unrealistically long time. Hence, this model is unusable for the real-time traffic assignment applications. To make the mentioned EF model computationally tractable, we have used MATLAB Curve Fitting Toolbox™ to linearize it. The linearized model is as follows

$$EFNO_{X_i}^t = 0.0302 v_i^t + 0.7 \quad \text{(gm/km)}, \ \forall i \in \mathcal{C} \setminus \mathcal{C}_S, \ \forall t \in \mathcal{T}. \tag{3.39}$$

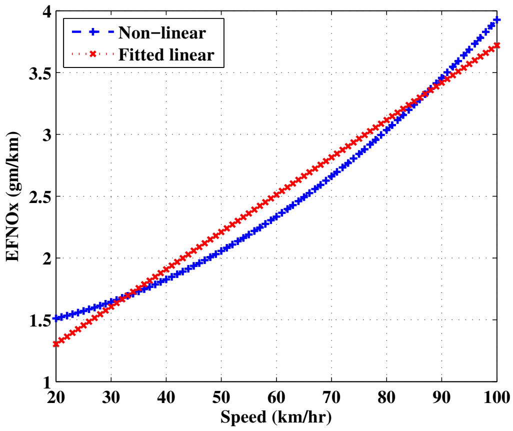


Figure 3.3. The non-linear Emission Factor (EF) model (in Equation (3.37)) is linearized and both of the non-linear model and the linear model (in Equation (3.39)) are plotted in the above figure. In the figure, we see that the linear model fits well with the original model. It is to be noted that the linearized model is within the 95% confidence bound of the original non-linear model.

The original non-linear model in Equation (3.37) and the linear approximation model in Equation (3.39) are plotted in Figure 3.3. The linearized model is within the 95% confidence bound of the original model. As a result, while using this model to measure EF, very less amount of error (i.e., maximum 5%) will occur (The MathWorks 2017). To determine a relationship between the DTA occupancy and the emission factor, we combine Equation (3.39) with the speed-occupancy relation in Equation (3.36) and get,

$$EFNO_{X_i}^t = 0.109\frac{V}{\tau}(1 - \frac{x_i^t}{N_i^t}) + 0.7 \quad (\text{gm/km}),\ \forall i \in \mathcal{C} \setminus \mathcal{C}_S,\ \forall t \in \mathcal{T}. \tag{3.40}$$

This is a linear formula to estimate emission for a given number of occupancy. However, if one uses the above formula as an objective function (i.e., to minimize total system-wide emission) in a DTA formulation, it would attain a solution with vehicle Holding-Back flows. Furthermore, it may cause unnecessary delays to the vehicles as maximizing occupancy within the network would minimize emission. To resolve this problem, we follow the similar approach as the NHB DTA-SC formulation presented in the previous section.

Our aim here is to derive an EB DTA-SC formulation that jointly minimizes total system-wide emission (TSE) and TSTT. Furthermore, we focus on deriving an emission framework that obtains Non-Holding-Back solution without imposing mixed-integer or non-linear programming problem. This advantage of the EB DTA-SC formulation enables it to be used to attain solution of any large size network within a very short time. Furthermore, the formulation has a direct relation with the CTM parameters and decision variables. This makes it very easy to implement. To formulate a combined traffic assignment and signal control formulation, we further include the presented signal control constraints (in Section 3.2.3.1) in the emission formulation. The complete EB DTA-SC formulation is given below.

$P_{EB\ DTA-SC}$:

minimize $$\mathcal{O}_{NO_X} = \sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C}} \left\{ 0.109 \frac{V}{\tau} (1 - \frac{x_i^t}{N_i^t}) + 0.7 \right\} \Theta_i, \ \Theta_i < \Theta_j, \tag{3.41}$$

subject to Equations (3.2)-(3.12), (3.24), (3.25)-(3.28), and

$$x_i^t = x_i^{t-1} + \sum_{k \in \Gamma^{-1}(i)} y_{ki}^{t-1}, \ \forall i \in \mathcal{C}_S, \ \forall k \in \Gamma^{-1}(i), \ \forall t \in \mathcal{T}. \tag{3.42}$$

In the objective function (3.41) of the EB DTA-SC formulation above, V is the free-flow speed, τ is the time-slot duration, x_i^t is the occupancy decision variable in the cell i at the time-slot t, N_i^t is the total occupancy capacity of that cell, and $j \in \Gamma(i)$. The objective function minimizes emission factor at the each of the cells at the each of the time-slots.

In the objective function (3.41) notice that the Θ_i incentive terms are multiplied with the EF at the corresponding cells. The value of Θ_i increases as the distance of the cell to the destination decreases. The Θ_i is minimum at the source cell and maximum at the sink cell. The Θ_i terms can be generated using the similar algorithm presented in Section 3.2.2 except the fact that the value of Θ_i should increase with the decreasing distance to the destination cell. As a result, these terms work as incentives to move vehicles to the destination cell and resolve the vehicle Holding-Back problem. This can be proven by following the same approach that is used to justify Proposition 3.2.3.

In the EB DTA-SC formulation, the sink cell is also included in the objective to determine emission based traffic assignment. This works as a pseudo incentive to maximize flow or maximize occupancy at the sink cell and moves the vehicles to the sink cell as quickly as possible. To update the sink cell occupancy, the flow-occupancy balance relation (3.42) is added to the formulation. At the destination cell $i \in \mathcal{C}_S$ one should set maximum occupancy capacity $N_i^t = \sum_{\forall t \in \mathcal{T}} \sum_{\forall k \in \mathcal{C}_R} d_k^t, \;\; i \in \mathcal{C}_S$ (i.e. total input demand). The presented EB DTA-SC objective function (3.41) minimizes emission factor by maximizing the ratio $\frac{x_i^t}{N_i^t}$ and the sink cell has highest incentive (i.e. maximum Θ_i) on maximizing this ratio. As a result, if one sets N_i^t equal to the total input demand, then the sooner all the vehicles reach the sink cell the better the objective value. Due to this property of the objective function, the EB DTA-SC formulation also maximizes throughput of the network and does not impose any unnecessary delay on the links while minimizing the emission factors at the each of the cells.

Once an EB DTA-SC solution becomes available, the total system-wide emission (TSE) can be computed by plugging in the optimal occupancies (x_i^t) to the Equation (3.38). In this equation notice that the sink cell is excluded. As a result, the emission at the sink cell is also excluded from the final estimation of the TSE.

3.3 Results and Discussions

In this section, firstly we discuss results for the proposed Non-Holding-Back SO-DTA (NHB SO-DTA) formulation. Afterward, solutions for various signal control models are analyzed and compared with the proposed SCRC model. Then emission results are discussed for different time-scales and control models.

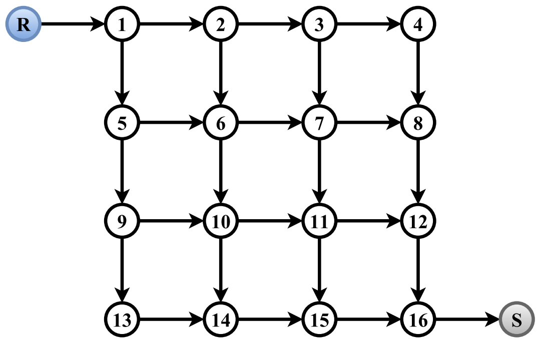


Figure 3.4. The grid network under investigation. This is the link and node representation of the implemented CTM network. In the figure, R is the source node, and S is the destination node. Node 6, 7, 10, and 11 are intersection nodes. Links from the nodes 13 to 15 and 4 to 12 are bottlenecks. The penalty terms for the CTM representation of this network were generated using the algorithm in Section 3.2.2.1.

3.3.1 Performance of the NHB SO-DTA Formulation

Consider the 4x4 grid network in Figure 3.4 where each of the links has a length of 2743.2 meters. As a result, each link can be represented by nine 20-second CTM cells when each of the cells has a length of 304.8 meters. All the cells have a maximum flow rate Q_i of 12 vehicles per slot, and jam density N_i of 36 veh per slot except the cells connecting the nodes 13 and 15, and the nodes 4 and 12. These cells have a flow capacity of 3 veh per slot and occupancy capacity of 9 veh per slot. The nodes 6, 7, 10, and 11 represent intersections. The generated demands at the source cell are {15, 30, 45, 60, 90, 120, and 180} vehicles arriving per time slot. The optimization was solved for a time horizon of 95 time-slots or 1900 seconds.

The SO-DTA results without considering signal control are presented in Table. 3.1. The HB SO-DTA objective in Equation (3.1) is equal to the Total System-wide Travel Time (TSTT) (Ziliaskopoulos 2000). In Table. 3.1, we see that all the three approaches attain the same TSTT value. It can also be seen in the Table. 3.1 that all the three frameworks have a comparable complexity regarding the number of variables and constraints, and yet the differences in the time required to solve them are significant. In particular, to obtain the Non-Holding-Back SO solution from the proposed framework, it is required to run the optimization twice (i.e., for SO solution

Table 3.1. Comparison of the objectives for the grid network.

Formulation	TSTT (veh-min)	Solving time (sec)	Variables	Constraints
HB SO-DTA	10282	421	44083	105168
Proposed NHB SO-DTA	10282	421+411	44083	105169
Feng Zhu NHB SO-DTA	10282	421+445+445[a]	44084	105170

[a]Firstly a very large value of M_L was chosen in Zhu's NHB SO-DTA framework. Then both the Zhu's NHB SO-DTA and Ziliaskopoulos' SO-DTA solutions were used to calculate an appropriate value of M_L (Zhu and Ukkusuri 2013).

we need to run the optimization formulation in Ziliaskopoulos (2000) then the proposed NHB SO-DTA formulation). Whereas, in the case of the approach presented in Zhu and Ukkusuri (2013) it was required to run the optimization at least three times (i.e., to determine correct parameters one needs to find solutions of the both of the formulations in Ziliaskopoulos (2000); Zhu and Ukkusuri (2013)). Furthermore, we have noticed that the latter might not find a solution with an incorrect choice of parameters (e.g., M_L). For example, when $M_L = 10^5$ was chosen for the grid network, then no solution was found.

3.3.2 CTM Time-scale and Traffic Signal Control

The accuracy of a DTA solution increases with the shorter time scale of CTM (Munoz et al. 2003; Sun and Bayen 2008). However, with shorter time scale, the length of a linear signal control cycle (i.e. in the CSDT model) also decreases. In contrast, the proposed SCRC model in this research allows us to set a realistic signal control cycle-length (e.g. 60s) while using a much shorter CTM time slot (e.g. 5s). For the sake of comparison, let us consider the 60s and 5s CTM time unit scenarios in Figure 3.5. In the figure, the cell 101 and 102 are intersection cells. The physical parameters of the two scenarios are summarized in Table. 3.2.

In this section, we are going to analyze the solutions of the CSDT and the SCRC signal control formulations and the impact of time-scale (i.e. time-slot duration) on the travel-time and evolution of traffic. The CSDT and SCRC signal control models

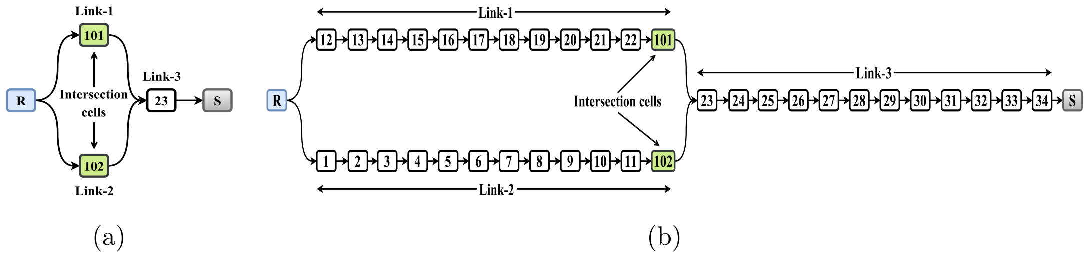


Figure 3.5. (a) The 60s cell network. (b) The 5s cell representation of the same network. Link and cell indexes are visible in the both of the figures. Both of the networks are solved for the optimization time-horizon of 20 minutes. CSDT signal control model is implemented in the 60s network. Whereas, the proposed SCRC model is embedded in the 5s network.

Table 3.2. Physical parameters of the 60s and the 5s CTM network.

	Parameters				
CTM network	Cell index	Q_i (veh per slot)	N_i (veh per slot)	v_i (ms^{-1})	δ_i
60s CTM	101	36	108	15.24	0.5
	102	20	60	15.24	0.5
	23	48	96	15.24	0.5
	R	1000	1000	-	-
	S	-	408	-	-
5s CTM	Cells in link-1	3	9	15.24	0.5
	Cells in link-2	1.67	5	15.24	0.5
	Cells in link-3	4	8	15.24	0.5
	R	1000	1000	-	-
	S	-	408	-	-

are separately used in the formulation. This allows us to compare the solutions of the mentioned control models and analyze their time-scale issues. Here we are interested in showing the difference between the 60s CTM with CSDT control model and the 5s CTM with SCRC control model when both of the control models have 60s cycle length (CL). Three different cases are studied that are given below.

- Case(1): The 60s CTM network in Figure 3.5(a) is solved by combining the CSDT signal control model (60s CL) with the NHB SO-DTA formulation and the input demands $\{d_R^1 = 12, d_R^2 = 24, d_R^3 = 36, d_R^4 = 48, d_R^5 = 96, d_R^6 = 192\}$.

- Case(2): The 5s CTM network in Figure 3.5(b) is solved using the NHB DTA-SC formulation in $P1$ (in Sub-section 3.2.4) with cycle-length of 60s. Please note, SCRC is the signal control model in this case. Like the case(1), the same demands $\{12, 24, 36, 48, 96, 192\}$ are uniformly distributed over the very first-six-minutes of the optimization time horizon. For example, the 12 vehicles demand is distributed such that 1 vehicle enters the network at the each of the 5s time-slots with time-slot index 1 to 12.

- Case(3): Same as case(2) except the fact that the demand is non-uniformly distributed as $\{d_R^1 = 12, d_R^{13} = 24, d_R^{25} = 36, d_R^{37} = 48, d_R^{49} = 96, d_R^{61} = 192\}$ (e.g. all the 12 vehicles are entering at the time-slot index 1 of the 5s time-slots).

It can be seen in the Table 3.3 that the CSDT model with the 60s CTM cell attains longer TSTT. This is because the optimization was done based on the average demand arrival at the source cell over the 60s period. Whereas, in the 5s uniform demand case, the optimization can utilize finer details of the demand arrival pattern. This allows the optimization to maximize the throughput with better knowledge of traffic propagation while reducing the TSTT. In the table, we see that the 5s uniform demand case attains 17.5% less TSTT. As a result, the 5s CTM cell case attains a better overall system-wide performance. The variation in demand has a significant impact

Table 3.3. Comparison of the different time-scales and control models

Scenarios	TSTT (veh-min)	Difference[a]
60s CTM CSDT[1]	2358	17.5%
5s CTM SCRC[2]	2008	-
5s CTM SCRC[3]	2188	9%

[1]Case(1); [2]case(2); [3]case(3); [a]computed with respect to 5s CTM uniform demand case(2).

on the optimal solution of the NHB DTA-SC problem. For the studied two 5s CTM scenarios, there is a significant 9% difference between the obtained TSTTs. This difference is due to the on-off demand pattern in the non-uniform case, which causes the demand to exceed the combined capacities available on link 1 and 2. As a result, these surplus vehicles have to wait at the source cell which then increases the total travel time. In the 60s case, there is no way one can distinguish between the uniform and non-uniform demands that differ within the coarse 60s slots. Therefore, a 5s case also captures demand arrival or departure pattern better.

Figure 3.6 depicts the travel time experienced by a vehicle traversing link 1, link 2, and the whole network, respectively. These travel times are computed from the cumulative arrival and departure flow curves of the corresponding links. In the figure, we observe that the 60s CTM cell case attains higher network-wide travel time. Whereas, there is a marginal difference between the two 5s cases. This is expected as the TSTTs of the 60s, and 5s cases are quite different. If one compares the 60s case with even shorter time-scale (e.g., 1s) then the difference in travel-time will increase further. From this result, we can conclude that the CTM time-scale has a notable impact on a DTA solution. Moreover, shorter time-slot duration should be chosen to attain optimal solution with acceptable accuracy. When this is the case, the proposed SCRC model provides a very convenient way to maintain a realistically large cycle length for the given very short time-scale.

It is also interesting to scrutinize the impact of time-scale on the vehicular traffic

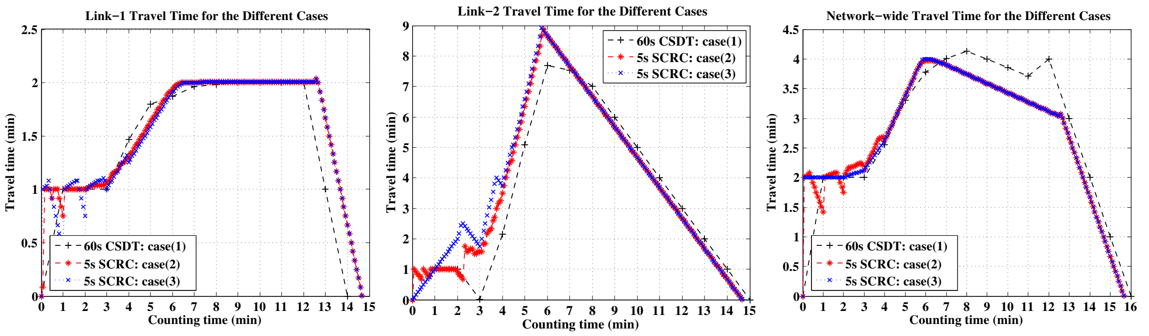


Figure 3.6. The figure shows travel-time experience by a vehicle while traversing the link 1, 2, and the network. Travel-times for all the scenarios under investigation are plotted. In the figure, we see that travel-time changes with changing time-scale and varying input demands to the network. This result also clarifies why one should use shorter time-scale.

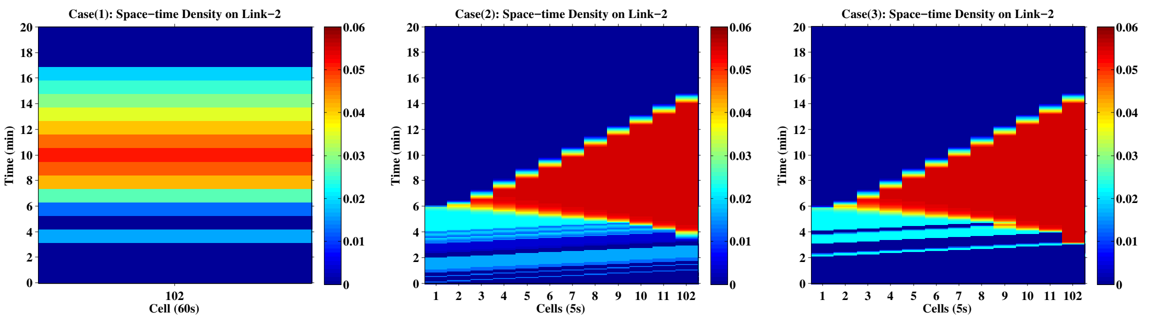


Figure 3.7. Distribution of space-time density over the link-2 for the different scenarios. In the figure, it is clearly noticeable that the both of the 5s CTM cases more accurately capture location of the congestion. Whereas, the 60s case provides only an average measure of density.

propagation and optimal traffic signal control settings. Space-time density profiles on link 2 for both the 60s and 5s scenarios are presented in Figure 3.7. In the figure, we observe that the density profiles are quite different for the scenarios under investigation, especially, at the intersection cell 102. This forces the optimization to apply different traffic signal control green splits for the different cases. In particular, the 60s CTM network loading model only provides an average picture of the traffic propagation. In this case, we see in the figure that the link is congested from the 8^{th} minute till the 14^{th} minute. Though this information can be used to determine which link is congested, it doesn't provide any particular location (e.g., within 76.2m range) of the queue formation or congestion. Whereas, more details about traffic propagation can be seen in the 5s CTM cases where the locations and dynamics of the queue are visible. As a result, shorter CTM time-scale helps us to locate

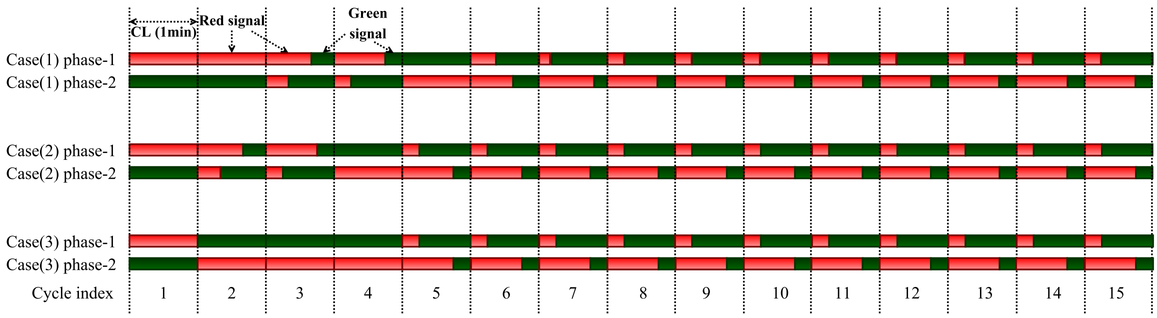


Figure 3.8. Figure shows first 15 optimal traffic signal timing attained for the case(1), case(2) and the case(3). In case(1), CSDT was used as the signal control model. Whereas, SCRC was implemented in the case(2) and the case(3). In the all of the cases, the signal control cycle-length (CL) is 60s. Though the case(2) and the case(3) time-slot duration is 5s, the SCRC model enables us to set the CL as 1 minute (min).

congestion on a particular part of a link. Furthermore, finer traffic dynamic (e.g., queue spillback) information can be attained using shorter time-scale. As a result, when the purpose of a DTA problem is to find out network bottlenecks, the shorter time-scale (e.g., 5s) must be used. Again, the proposed SCRC model would provide us a strong foundation to append traffic signal control in the shorter time-scale DTA scenario.

We have implemented different control methods in the proposed NHB SO-DTA formulation and have created three different cases that were discussed earlier. The optimal traffic signal control timing for all these cases are presented in Figure 3.8. Green-times for both the phase-1 (cell 101) and phase-2 (cell 102) are presented in the figure for these cases. In the space-time density Figure 3.7 we see that congestion starts developing at link-2 after the 6^{th} minute in case(1). In the case(2), the intersection cell 102 becomes congested after the 4^{th} minute. Whereas, in case(3) this cell gets congested after the 3^{rd} minute. As a result, during these periods of time when the networks are under-saturated the differences among the signal settings are significant. However, when the networks are congested (e.g., 5^{th} minute in case(1) and case(2)) the signal control green-proportions are determined by the available occupancy space at the downstream cell 23. For instance, the maximum occupancy capacity of the cell 23 is 96 vehicles over the 60s time. During the congested period, 32 vehicles are already occupying cell 23. As a result, due to the constraint (3.11) another 32 vehicles can enter to cell 23. The cell 101 has saturation flow capacity of 36 vehicles, and the cell 102 has that of 20 vehicles over the 60s time. Now if one allocates 45s (i.e., serves 24 vehicles) to phase-1 and 15s to phase-2 (i.e., serves eight vehicles) then they add up to the remaining accommodation capacity of 32. Due to this fact, in the congested network condition (e.g., 7^{th} to 15^{th} minute) the attained signal settings are similar in all these three cases. It is also noticeable in the figure that the sum of the green-times within a signal control cycle for the phase-1 and phase-2 adds up to the signal control cycle-length. This happens due to the constraint (3.28). It is to be noted that the proposed SCRC model has less computational complexity compared to the CSDT model. This is evident especially when the SCRC model is implemented in a shorter time scale (e.g., 5s) network with a CL is a multiple of time slots. As the SCRC signal control constraints and variables are applied only over the cycle-length (e.g., twelve 5s slots or one 60s) rather than each of the time-slots (e.g., 5s), the SCRC scheme imposes less complexity. In the next section, an example is provided to show this gain regarding complexity.

3.3.2.1 Signal Control Performance and Complexity

In this section we study the complexities of the NHB DTA-SC framework defined in sub-section 3.2.4 with different control methods using the grid network in Figure 3.4 but without the nodes 4 and 13 and their associated links. These links were removed to make the problem smaller so that we can obtain a solution using the MISC.

Each of the links in the considered network consists of nine 20s CTM cells, and there are four intersections in the network (i.e., node 6, 7, 10 and 11). For this study, only two phases are considered that are North-South and West-East traffic movements. The signal controls considered here are the CSDT (20s CTM, 20s cycle length), MISC (20s CTM, 20s cycle length), and SCRC (20s CTM, 60s cycle length)

Table 3.4. Comparison Between the Signal Control Models

Control	TSTT (veh-min)	Solving time	Variables	Constraints	Complexity*
CSDT[1]	53970	911 (sec)	48723	115083	1.6%
SCRC[2]	53970	812 (sec)	47443	113803	-
MISC[3]	53987	1857 (sec)	50643	118924	5.2%

*Higher complexity compared to the SCRC method. Computed from the differences in sum of variables and constraints; [1]In the 20s CTM network with 20s cycle-length; [2]in the 20s CTM network with 60s cycle-length; [3]in the 20s CTM network with 20s cycle-length

defined in sub-section 3.2.3. The demand to the network is {15, 30, 45, 60, 90, 120, 150, 180, and 200} vehicles arriving at the discrete time slots 1 to 9, respectively. It can be seen in Table. 3.4 that the SCRC and CSDT models have the same TSTT while there is only a minor difference of 1.6% in complexity. Note that the number of variables and constraints in SCRC model are reduced as the signal control variables, and constraints are created only at each cycle (of multiple time slots) instead of at each time slot. As a result, the solution time is also reduced. For a larger network, this complexity gain is expected to improve further. Clearly, the MISC model has the highest complexity, and due to the binary nature of the green split variables, it took more than twice the time required by the SCRC model to attain the optimal solution.

3.3.3 Accuracy of the Proposed EB DTA-SC Formulation

In Section 3.2.6 we have presented an Emission Based Dynamic Traffic Assignment with Signal Control (EB DTA-SC) formulation. The original non-linear Emission Factor (EF) model in Equation (3.37) is linearized to Equation (3.39). To show the accuracy of the approximate linear EF model we investigate the following three scenarios.

- Scenario(a): The 60s network (in Figure 5(a)) is solved using the CSDT signal control model (60s CL) embedded in the EB DTA formulation with demands $\{d_R^1 = 12, d_R^2 = 24, d_R^3 = 36, d_R^4 = 48, d_R^5 = 96, d_R^6 = 192\}$.

Table 3.5. Accuracy of the proposed emission based formulation

Formulation	TSE (gm) (approximate EF model)[†]	TSE (gm) (original EF model)[‡]	Error[d]
60s EB DTA-SC[a]	4421	4280	3.3%
5s EB DTA-SC[b]	3341	3336.4	0.13%
5s EB DTA-SC[c]	3671.5	3663	0.24%

[a]Scenario(a);[b]scenario(b); [c]scenario(c);[†]in Equation (3.39);[‡]in Equation (3.37); [d]with respect to the corresponding original EF model case.

- Scenario(b): The 5s CTM network in Figure 5(b) is solved using the proposed EB DTA-SC formulation in $P_{EB\ DTA-SC}$. In this case, SCRC is the embedded signal control model with cycle-length of 60s. The same demands $\{12, 24, 36, 48, 96, 192\}$ are uniformly distributed over the $\{1, 2, 3, 4, 5, 6\}^{th}$ minute of the optimization time horizon, respectively.

- Scenario(c): Same as scenario(b) but the demand is distributed as $\{d_R^1 = 12, d_R^{13} = 24, d_R^{25} = 36, d_R^{37} = 48, d_R^{49} = 96, d_R^{61} = 192\}$ over the 5s time-slots $\{1, 13, 25, 37, 49, 61\}$.

The emission measures attained using the approximate EF model and the original non-linear models for these three scenarios are presented in Table 3.5. In the table, we observe that in the 60s case the difference between the measured total system-wide emission (TSE) using the approximate model and the original model is 3.3%. On the other hand, in the shorter time-scale scenarios (i.e. 5s) the error reduces below 0.25%. In all these cases, the error between the original model and the approximate linear model is below 5%. This validates the accuracy of the approximate linearized model. As it is within the 95% confidence bound of the original model, the error would always remain within 0% to 5% (The MathWorks (2017)).

3.3.4 Improved Vehicle Discharged Emission

Vehicle discharged emission is directly connected to speed and density. Likewise, the speed and density of the vehicles are controlled by the DTA, and the time-scale has a notable impact on the DTA. As a result, it would be interesting to see the effect of time scale on emission. To analyze this impact on emission, we investigate the following three new scenarios along with the scenarios presented in Section 3.3.3.

- Scenario(a1): The 60s CTM network in Figure 5(a) is solved using the CSDT signal control model (60s CL) embedded in the HB SO-DTA formulation P_Z presented in Section 3.2.1 with the same demand as scenario(a).
- Scenario(b1): The 5s CTM network in Figure 5(b) is solved using the HB SO-DTA formulation when SCRC is the embedded signal control model with cycle-length of 60s. The input demand is same as that of the scenario(b) in the previous section.
- Scenario(c1): The same formulation is used as scenario(b1) with the demand of scenario(c) discussed in the previous section.

The numerical results for all these cases are presented in Table 3.6. In the table, we can see that the EB DTA-SC formulation attains less total system-wide emission (TSE) (i.e., E_{NO_X}) for the same total system-wide travel-time (TSTT). This confirms the fact that the proposed EB DTA-SC formulation attains traffic assignment solutions by minimizing emission without causing extra delay to the vehicles. In the table, it is also noticeable that there is 26% difference between the 60s scenario(a1) and the 5s scenario(b1) TSE. Similarly, the difference between the 60s and 5s case E_{NO_X} measured by the EB DTA-SC formulation is 32%. The 5s CTM has higher resolution and captures the traffic dynamics more accurately which contributes to this difference in NO_X emission. A similar result was found in the previous section while analyzing the TSTT. This result verifies that the time-scale has a significant impact on the estimation process of vehicle discharged emission.

Table 3.6. Travel-time and total system-wide emission for the six scenarios.

Scenarios	TSTT	E_{NO_X} (gm)	Formulation E_{NO_X} Difference†	Time-scale E_{NO_X} Difference‡
60s HB SO-DTA SC[a1]	2358	4778	8%	26%
60s EB DTA-SC[a]	2358	4421	-	32%
5s HB SO-DTA SC[b1]	2008	3783	13%	-
5s EB DTA-SC[b]	2008	3341	-	-
5s HB SO-DTA SC[c1]	2188	4085	11.3%	-
5s EB DTA-SC[c]	2188	3671.5	-	-

[a1]Scenario(a1); [a]scenario(a); [b1]scenario(b1); [b]scenario(b); [c1]scenario(c1); [c]scenario(c); †compared to the same time-scale of the EB DTA-SC scenario; ‡scenario(a1) is compared to the scenario(b1) and scenario(a) is compared to the scenario(b).

3.4 Chapter Summary

The conventional System Optimal Dynamic Traffic Assignment (SO-DTA) formulation suffers from the known vehicle Holding-Back (HB) problem. The SO-DTA formulation with HB problem imposes unrealistic traffic propagation over the links. This makes the HB SO-DTA solution unusable in the real-life traffic assignment applications. Since signal control directly influences the flow and density of traffic on the roads, while studying DTA, signal control must be added to the formulation. However, most of the existing signal control models are mixed-integer or nonlinear and have very high complexity. Even if there are few linear signal control models, while using them one can only set signal control cycle-length (CL) equal to the DTA discrete time-interval. As the accuracy of the solution increases with shorter time-scale, this restriction of time-scale trades accuracy for a realistic cycle-length (e.g., 60s).

To address the problems mentioned above, we propose a novel linear and continuous NHB DTA-SC framework. The proposed framework combines signal control and traffic assignment on the same platform and attains NHB SO solution. Moreover, this formulation is straightforward to implement and does not require any recursive parameter selection. Also, the SCRC model embedded in the formulation enables us to set any cycle-length (multiple time-slots) for any time-scale. As a result, it equips us to strike a right balance between accuracy and realistic control policy. We have found that the shorter time-scale used in the proposed NHB DTA-SC framework captures traffic propagation characteristics better. Results show that there is up to 17.5% difference in TSTT between the scenarios using the 5s and the 60s time-slots.

We have also proposed a linear NHB Emission-Based Dynamic Traffic Assignment with Signal Control (EB DTA-SC) formulation. We have estimated NO_X emission using the proposed EB DTA-SC formulation. Results confirm the fact that the proposed EB DTA-SC formulation minimizes vehicle discharged emission without imposing unnecessary delay to the road users. Furthermore, the estimated emission using the 5s and the 60s time-slots differed up to 32%. This indicates that the time-scale employed in the DTA formulation has a notable impact on the emission estimation.

Chapter 4

Structures of the Optimal Solutions

4.1 Introduction

The Non-Holding-Back Dynamic Traffic Assignment (NHB SO-DTA) closely follows real-life traffic propagation characteristics. Furthermore, an NHB solution is much easier to implement and less costly compared to its HB counterpart (Shen, Nie, and Zhang 2007). As a result, understanding the optimal solution structures of the NHB SO-DTA would equip us better to determine more advanced traffic management strategies such as automated intersection control, ramp metering, vehicle platooning or emergency evacuation planning.

The major focus of this chapter is to analyze optimal solution structures of both of the Non-Holding-Back (NHB) traffic assignment and the signal control formulation. The NHB SO-DTA formulation was presented in the previous chapter, and it was proven that the formulation completely resolves known vehicle Holding-Back (HB) problem. In this chapter, firstly the optimal solution structures of the mentioned formulation are presented and rigorously proven. These structures are related to both of the optimal flow and occupancy decision variables of the widely used Cell Transmission Model (CTM) Dynamic Network Loading (DNL) model. An NHB intersection-delay-minimization (NHB IDM) formulation is also presented. We compare the solution of NHB IDM with the proposed NHB SO-DTA formulation and show that the latter approach also minimizes intersection delay. The contributions related to the structures of the optimal solution of the NHB SO-DTA formulation are enlisted below.

- It is demonstrated that by keeping the optimal flows unchanged the penalty based NHB SO-DTA formulation can attain multiple optimal solutions by interchanging the occupancies within the same penalty level.
- It is also verified that for a single path network the NHB SO-DTA formulation attains a unique solution.
- It is proven that the aggregated optimal flow of the formulation within a penalty level and at a time-slot must have to be equal and maximum.
- It is also shown that the NHB SO-DTA formulation minimizes overall network-wide delay.
- Finally, it is justified that the proposed NHB SO-DTA formulation minimizes aggregated intersection delay.

The Non-Holding-Back solution captures real-life traffic propagation properties better, and a meaningful study of the signal control can be carried out within such a formulation. Thus, in the second part of this chapter, the optimal signal control structures in a Non-Holding-Back (NHB) system optimal DTA framework are studied for different traffic conditions (e.g., under/over-saturated). A Lagrange multiplier method is used to derive the solution structures. Lagrange multiplier method is the most widely accepted approach to analyze optimality conditions of an optimization formulation. To the best of author's knowledge, this is the first approach using such a method to study the optimal signal control settings in a DTA framework. The key contributions related to the signal control structures are:

- develop a decomposition of the NHB DTA-SC problem using Lagrangian dual decomposition method and formulate three separate sub-problems of *occupancy minimization*, *flow maximization*, and *signal control*;
- mathematically derive the Karush-Kuhn-Tucker (KKT) optimality conditions of the decomposed signal control sub-problem in over-saturated, under-saturated, and queue spillback traffic conditions;
- show and prove the following *locally optimal* structures of the signal control in the proposed NHB DTA-SC framework
  - in an over-saturated condition the optimal control follows the "equisaturated" control policy of Webster (1958);
  - in an under-saturated condition the optimal control follows the "proportional" control policy of Le et al. (2015);
  - and in a queue spillback scenario the optimal control follows a control policy such that the sum of the green-splits of the different phase movements is proportional to the available capacity downstream.

The remainder of the chapter is organized as follows. In Section 4.2 the structures and analysis of the optimal solution of the proposed NHB SO-DTA formulation are presented. For comparison purposes, an Intersection Delay Minimization formulation is also outlined in this section. Section 4.3 presents numerical results related to the NHB SO-DTA optimal solution structures. The NHB DTA-SC formulation is presented in Section 4.4. This formulation is decomposed, and all the theoretical optimal signal control structures in the NHB DTA-SC framework are outlined in Section 4.5. Section 4.6 shows the numerical convergence of all these theoretical structures. Finally, Section 4.7 concludes the chapter by summarizing the key outcomes of this study.

4.2 Structures of the NHB SO-DTA Formulation

In this section, we discuss solution structures of the proposed NHB SO-DTA formulation. The complete NHB SO-DTA formulation discussed in the previous chapter is given below for convenience.

$P_{NHB\ SO\text{-}DTA}$:

$$\text{maximize} \quad \mathcal{O}_{NHB} = \sum_{\forall t \in \mathcal{T}} \sum_{\forall ij \in \mathcal{E}_S} y_{ij}^t - \sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C} \setminus \mathcal{C}_S} \Psi_i \cdot x_i^t \tag{4.1}$$

subject to Equations (3.2)-(3.12), and

$$\sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C} \setminus \mathcal{C}_S} x_i^t = \tilde{\mathcal{O}}_Z, \tag{4.2}$$

where $\tilde{\mathcal{O}}_Z$ is attained using Equations (3.1)-(3.12),

and the non-negativity constraints $y_{ij}^t \geq 0,\ x_i^t \geq 0,\ g_i^{c,\phi_p} \geq 0$.

To examine the solution of the NHB formulation, it is important to understand penalty levels and relative distance of the cells from the destination cell. A penalty level l is determined by the algorithm presented in Section 3.2.2.1. The algorithm is based on the fact that the nearest cells to the destination cell have a lower penalty on the occupancies. Whereas, the penalty increases with the increasing distance from the destination cell. Cells that have the same penalty (i.e., same distance from the sink cell) belong to the same penalty level, and we define it as $\mathcal{C}^l$. We prove that $P_{NHB\ SO\text{-}DTA}$ may attain multiple solutions because occupancies are interchangeable at the same penalty level only and at different penalty levels the solutions must be same. In Equation (4.1) several cells have same penalty term (Ψ_i) if they are at the same distance from the destination cell. Such cells are said to be at the same penalty level.

The major contributions of this chapter are the optimal-solution structures of the NHB SO-DTA formulation presented in $P_{NHB\ SO\text{-}DTA}$. The structures are outlined below in the form of lemmas, theorems, and propositions. The physical interpretations and applications of such structures are also discussed.

To clearly indicate penalty levels at the each of the occupancy and flow variables, we define the occupancy x_i^t of cell i at penalty level l as x_i^{lt}, where i is the cell index and l is the penalty level indicator. Likewise, we define flow y_{ij}^t as y_{ij}^{lt} which is the flow going from cell i at penalty level l to cell j at penalty level l' where $\forall i \in \mathcal{C}^l$.

4.2.1 Structures of Optimal Occupancy

Lemma 4.2.1. *At any time-slot t_0 and at any penalty level $\forall l \in \mathcal{L}$, if we consider two different solutions $(\bar{x}, \bar{y})$ and $(x^\star, y^\star)$ of the NHB SO-DTA formulation ($P_{NHB\ SO\text{-}DTA}$), then these solutions have property due to the Non-Holding-Back nature such that*

$$\sum_{\forall l \in \mathcal{L}} \sum_{\forall i \in \mathcal{C}^l} x_i^{\star l t_0} = \sum_{\forall l \in \mathcal{L}} \sum_{\forall i \in \mathcal{C}^l} \bar{x}_i^{l t_0}, \quad \sum_{\forall t \le t_0} \sum_{\forall j \in \mathcal{C}_S} \bar{y}_{ij}^{l t_0} = \sum_{\forall t \le t_0} \sum_{\forall j \in \mathcal{C}_S} y_{ij}^{\star l t_0}, \ \forall \mathcal{C}_S \notin \mathcal{L}. \tag{4.3}$$

The above two properties imply that the total occupancy in the network at a time-slot t_0 is same for any optimal solution of $P_{NHB\ SO\text{-}DTA}$. In another way, total throughput of the network till t_0 at a sink cell is equal for any solution of the formulation. Otherwise, one can always increase the flow at the sink-cells (decrease occupancy in the network) and can attain better solution than $(x^\star, y^\star)$ and $(x^\star, y^\star)$ becomes sub-optimal.

Theorem 4.2.2. *The solutions of the formulation $P_{NHB\ SO\text{-}DTA}$ have the same aggregated occupancy at any penalty level l at time t_0 which is*

$$\sum_{\forall i \in \mathcal{C}^l} \bar{x}_i^{l t_0} = \sum_{\forall i \in \mathcal{C}^l} x_i^{\star l t_0}. \tag{4.4}$$

Proof. For this proof we consider two solutions $(\bar{x}, \bar{y})$ and $(x^\star, y^\star)$ of the formulation $P_{NHB\ SO\text{-}DTA}$. Their objective values are denoted by $\bar{\mathcal{O}}_1$ and $\mathcal{O}_1^\star$, respectively. We assume at a time interval t_0 and at a penalty level $\mathcal{C}^l$ the solutions of the proposed NHB formulations are such that,

$$\sum_{\forall i \in \mathcal{C}^l} \bar{x}_i^{l t_0} > \sum_{\forall i \in \mathcal{C}^l} x_i^{\star l t_0}. \tag{4.5}$$

In the above inequality, x_i^{lt}'s are optimal occupancies and $\mathcal{C}^l$ are the cells at penalty level l where i is the cell index. From the Lemma 4.2.1 above one can deduce that the remaining amount of traffic is somewhere in the network and we can define,

$$\Delta^l = \sum_{\forall i \in \mathcal{C}^l} \bar{x}_i^{lt_0} - \sum_{\forall i \in \mathcal{C}^l} x_i^{\star lt_0}. \tag{4.6}$$

In $(\bar{x}, \bar{y})$ and $(x^\star, y^\star)$ we assume this quantity of traffic is at another cell which has penalty level $l' \neq l$. At these two penalty levels, for the two solutions $(\bar{x}, \bar{y})$ and $(x^\star, y^\star)$ we can write,

$$\Delta^l + \sum_{\forall i \in \mathcal{C}^l} x_i^{\star lt_0} = \sum_{\forall i \in \mathcal{C}^l} \bar{x}_i^{lt_0}. \tag{4.7}$$

Since the solution $(x^\star, y^\star)$ is attained from the proposed NHB formulation, this difference in occupancy is somewhere within the network $l' \neq l$. Due to this fact, following can be written at penalty level l'.

$$\sum_{\forall j \in \mathcal{C}^{l'}} x_j^{\star l't_0} - \Delta^l = \sum_{\forall j \in \mathcal{C}^{l'}} \bar{x}_j^{l't_0}, \ \forall i \in \Gamma^{-1}(j). \tag{4.8}$$

It is known that penalty levels $\Psi_k > \Psi_i > \Psi_j, \forall k \in \Gamma^{-1}(i), \forall i \in \Gamma^{-1}(j)$. Now if we take difference between $\mathcal{O}_1^\star$ and $\bar{O}_1$ and substituting Equations (4.7) and (4.8) we get,

$$\begin{aligned}
\mathcal{O}_1^\star - \bar{\mathcal{O}}_1 &= \sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C} \setminus \mathcal{C}_S} \Psi_i \cdot \bar{x}_i^t - \sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C} \setminus \mathcal{C}_S} \Psi_i \cdot x_i^{\star t} \\
&= \Psi_i \left(\sum_{\forall i \in \mathcal{C}^l} \bar{x}_i^{lt_0} - \sum_{\forall i \in \mathcal{C}^l} x_i^{\star lt_0} \right) + \Psi_j \left(\sum_{\forall j \in \mathcal{C}^{l'}} \bar{x}_j^{l't_0} - \sum_{\forall j \in \mathcal{C}^{l'}} x_j^{\star l't_0} \right) \\
&= \Psi_i \left(\Delta_i^l + \sum_{\forall i \in \mathcal{C}^l} x_i^{\star lt_0} - \sum_{\forall i \in \mathcal{C}^l} x_i^{\star lt_0} \right) + \Psi_j \left(\sum_{\forall j \in \mathcal{C}^{l'}} x_j^{\star l't_0} - \Delta_i^l - \sum_{\forall j \in \mathcal{C}^{l'}} x_j^{\star l't_0} \right) \\
&= (\Psi_i - \Psi_j) \cdot \Delta_i^l \neq 0.
\end{aligned}$$

This contradicts with the fact that both the $\mathcal{O}_1^\star$ and $\bar{O}_1$ are optimal solutions of $P_{NHB\ SO\text{-}DTA}$. As a result, to become optimal solution Δ_i^l has to be zero. Hence, the optimal solutions of $P_{NHB\ SO\text{-}DTA}$ can be permuted within the same penalty levels only. □

In real life traffic scenarios, it is evident that the vehicles will move forward as long as they are not blocked by congestion or link failures. In NHB SO-DTA framework the penalty on the objective function decreases towards the destination. This property forces the vehicles to move closer to the destination at a sooner possible time without changing the overall Total System-wide Travel Time (TSTT). As a result, unlike HB SO-DTA solution this solution is more representative of the real-life traffic propagation.

Moreover, for a single path network each of the penalty level contains only one cell. In this case, according to the Theorem 4.2.2 the NHB SO-DTA formulation will attain unique solution. This has been shown in the following lemma.

Lemma 4.2.3. *For a single path network $P_{NHB\ SO\text{-}DTA}$ attains unique solution.*

Proof. The single path network has unique penalty levels for each cells which is $\nexists \Psi_j | \Psi_j = \Psi_i, \forall i \in \mathcal{C} \setminus \mathcal{C}_S \cup \mathcal{C}_D \cup \mathcal{C}_M,\ l \in \mathcal{C}^l$. Then the Theorem 4.2.2 implies that,

$$\bar{x}_i^{lt} = x_i^{\star lt},\ \forall t \in \mathcal{T}. \tag{4.9}$$

As a result, the proposed formulation attains unique solution for a single path network. □

4.2.2 Structure of the Optimal Flow

In the next theorem we present an optimal solution structure of $P_{NHB\ SO\text{-}DTA}$ related to the optimal flows at the each of the penalty levels.

Lemma 4.2.4. *At any time-slot t_0 and at a penalty level l, $P_{NHB\ SO\text{-}DTA}$ attains maximum possible flow, which is, if we consider $(x^\star, y^\star)$ as the solution of $P_{NHB\ SO\text{-}DTA}$ and $(\bar{x}, \bar{y})$ as another NHB solution, then following holds*

$$\sum_{\forall i \in \mathcal{C}^l} \bar{y}_{ij}^{lt_0} \leq \sum_{\forall i \in \mathcal{C}^l} y_{ij}^{\star lt_0}. \tag{4.10}$$

Proof. We know that the formulation in $P_{NHB\ SO\text{-}DTA}$ is NHB. As a result, for a specific cell and at a particular time-slot t_0, one of the inequality constraints (3.2)-(3.12) must satisfy with an equality constraint. Otherwise, it contradicts with the NHB property of the formulation. For that cell, if one of these inequalities is met by an equality then the outflow of that cell is the maximum attainable flow at that time-slot. This proves Lemma 4.2.4. □

From the above lemma, it is clear that the NHB SO-DTA formulation maximizes flow at the each of the penalty level, which eventually maximizes overall network-wide throughput. This is a very important structure and in the following theorems we will show how this property helps $P_{NHB\ SO\text{-}DTA}$ to minimize network-wide delay.

4.2.3 Structures of Delay in the Network

For the CTM networks, the definition of delay was presented by Lo (1999). Based on the definition, we formulate a NHB Delay Minimization (NHB DM) formulation. The formulation is presented below. This formulation locally minimizes delay at the each of the cell. To preserve Non-Holding-Back property, we add the optimal objective value of $P_{NHB\ SO\text{-}DTA}$ as a constraint (in Equation (4.12)).

$P_{NHB\ DM}$:

minimize $\quad \mathcal{O}_{DM} = \sum_{\forall l \in \mathcal{L}} \sum_{\forall i \in \mathcal{C}^l} \sum_{\forall t \in \mathcal{T}} (x_i^{lt} - y_{ij}^{lt}),\ \forall \mathcal{C}_S \notin \mathcal{L} \quad (4.11)$

subject to $\quad$ Equations (3.2)-(3.12),

$$\sum_{\forall t \in \mathcal{T}} \sum_{\forall l \in \Gamma^{-1}(j), \forall ij \in \mathcal{E}_S} y_{ij}^{lt} - \sum_{\forall l \in \mathcal{L}} \sum_{\forall i \in \mathcal{C}^l} \sum_{\forall t \in \mathcal{T}} \Psi_i \cdot x_i^{lt} = \mathcal{O}^{\star}_{NHB}. \quad (4.12)$$

The NHB DM formulation is presented here because we are going to compare the NHB SO-DTA solution with that of the NHB DM formulation.

Theorem 4.2.5. *The formulation in $P_{NHB\ SO\text{-}DTA}$ minimizes delay ($\mathcal{Q}^{lt_0}$) at the each of the penalty levels l at any time-slot t_0. For the solution $(\bar{x}, \bar{y})$ of $P_{NHB\ SO\text{-}DTA}$*

and $(x^\star, y^\star)$ *of* $P_{NHB\ DM}$ *this property can be written as*

$$\mathcal{Q}^{lt_0} = \sum_{\forall i \in \mathcal{C}^l} (\bar{x}_i^{lt_0} - \bar{y}_{ij}^{lt_0}) = \sum_{\forall i \in \mathcal{C}^l} (x_i^{\star lt_0} - y_{ij}^{\star lt_0}),\ \forall i \in \mathcal{C} \setminus \mathcal{C}_S. \tag{4.13}$$

Proof. The actual residual queue or delay at a penalty level l at the time-slot t_0 can be defined in terms of occupancy and flow as,

$$\mathcal{Q}^{lt_0} = \sum_{\forall i \in \mathcal{C}^l} (x_i^{lt_0} - y_{ij}^{lt_0}). \tag{4.14}$$

From the Lemma 4.2.4 we know that $P_{NHB\ SO\text{-}DTA}$ always attains maximum flow at the each of the penalty levels at any time-slot. Same property persists in $P_{NHB\ DM}$ as we deploy the objective value of $P_{NHB\ SO\text{-}DTA}$ as a constraint in this formulation. Hence, one can write

$$\sum_{\forall i \in \mathcal{C}^l} \bar{y}_{ij}^{lt_0} = \sum_{\forall i \in \mathcal{C}^l} y_{ij}^{\star lt_0}. \tag{4.15}$$

Similarly, from the Theorem 4.2.2 we can deduce,

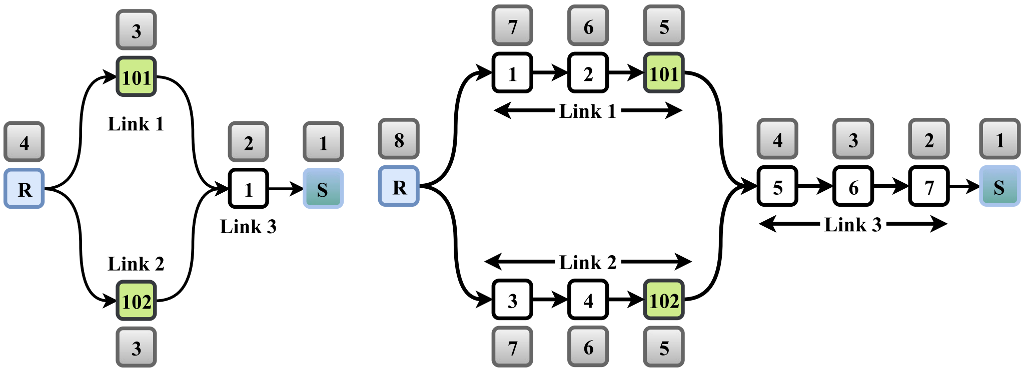
$$\sum_{\forall i \in \mathcal{C}^l} \bar{x}_i^{lt_0} = \sum_{\forall i \in \mathcal{C}^l} x_i^{\star lt_0}. \tag{4.16}$$

By subtracting Equation (4.15) from Equation (4.16) one gets Equation (4.13) and the Theorem 4.2.5 holds. □

Theorem 4.2.6. *The formulation in* $P_{NHB\ SO\text{-}DTA}$ *minimizes total system-wide delay. In other words,* $P_{NHB\ SO\text{-}DTA}$ *attains System Optimal (SO) solution. If* $(\bar{x}, \bar{y})$ *and* $(x^\star, y^\star)$ *are solutions of* $P_{NHB\ SO\text{-}DTA}$ *and* $P_{NHB\ DM}$ *respectively, then mathematically this structure can be written as*

$$\sum_{\forall l \in \mathcal{L}} \sum_{\forall i \in \mathcal{C}^l} \sum_{\forall t \in \mathcal{T}} (\bar{x}_i^{lt} - \bar{y}_{ij}^{lt}) = \sum_{\forall l \in \mathcal{L}} \sum_{\forall i \in \mathcal{C}^l} \sum_{\forall t \in \mathcal{T}} (x_i^{\star lt} - y_{ij}^{\star lt}),\ \forall \mathcal{C}_S \notin \mathcal{L}. \tag{4.17}$$

Proof. To prove that $P_{NHB\ SO\text{-}DTA}$ minimizes overall delay in the network, let us recall the Theorem 4.2.5. If we use this theorem to find delay throughout the whole network, Theorem 4.2.6 immediately follows. □



(a) The 60s CTM network.

(b) The 20s CTM network.

Figure 4.1. Figures show 60s and 20s cell representation of the same network. The numbers in the shaded boxes are the penalty levels (Ψ_i) for the corresponding cells. In the figure we see that the penalty levels are decreasing with increasing distance from the source cell.

Lemma 4.2.7. *The formulation in $P_{NHB\ SO\text{-}DTA}$ minimizes aggregated intersection delay such that*

$$\sum_{\forall i \in \mathcal{C}^l} (\bar{x}_i^{lt_0} - \bar{y}_{ij}^{lt_0}) = \sum_{\forall i \in \mathcal{C}^l} (x_i^{\star lt_0} - y_{ij}^{\star lt_0}),\ \forall i \in \mathcal{C}_I. \tag{4.18}$$

Proof. If we apply the Theorem 4.2.5 to measure aggregated intersection delay, then Lemma 4.2.7 is obtained.

□

4.3 NHB SO-DTA Numerical Solutions

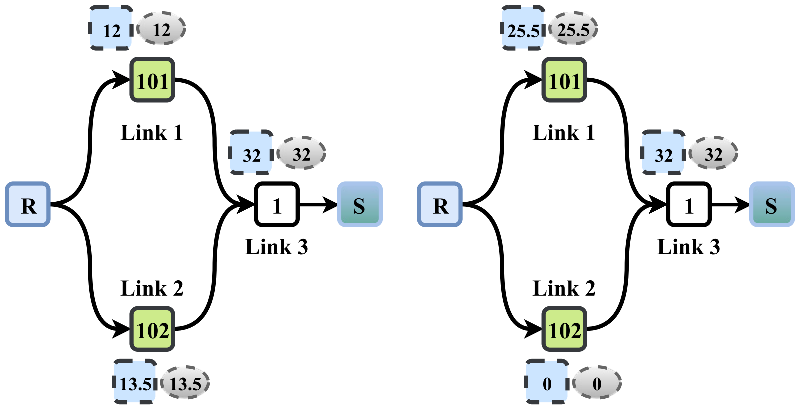
In this section, the results related to the discussed NHB traffic assignment and signal control structures are presented. In Figure 4.1 the cell 101 and 102 are intersection cells. The physical parameters of the 60s and 20s scenarios are summarized in Table. 4.1. Note that the network considered in Figure 4.1 is non-symmetric where link 1 (cell 1, 2, 101) has a higher flow and occupancy capacity than link 2 (cell 3, 4, 102). Also link 3 (cell 5, 6, 7) has less capacity than the combined capacity of links 1 and 2.

Table 4.1. The physical parameters of 60s and 20s CTM network.

	Parameters				
CTM network	Cell index	Q_i (veh per slot)	N_i (veh per slot)	v_i (ms^{-1})	δ_i
60s CTM	101	36	108	15.24	0.5
	102	20	60	15.24	0.5
	1	48	96	15.24	0.5
20s CTM	1, 2, 101	12	36	15.24	0.5
	3, 4, 102	6.67	20	15.24	0.5
	5, 6, 7	16	32	15.24	0.5

4.3.1 Interchangeable Occupancy

This structure was discussed in Theorem 4.2.2. For this structure let's consider the 60s CTM network in Figure 1(a). The generated demands at the source cell are $d_R^1 = 90$, $d_R^2 = 180$, and $d_R^3 = 270$ vehicles. The optimization was run for 22 time-slots or 1320 seconds (s). Firstly, we solve the SO-DTA optimization problem to obtain Total System-wide Travel Time (TSTT). This value is used as $\tilde{\mathcal{O}_Z}$ in Equation (4.2) to obtain system optimal (SO) solution in the NHB SO-DTA formulation. To show that the solution of the NHB SO-DTA formulation is permutable at the same penalty level, let's consider the optimal solution of the network in Figure 1(a) at time slot $t = 18$. The solution is presented in Figure 2(a). In the figure, we see that 12 vehicles are remaining at the cell 101 and cell 102 has occupancy of 13.5 vehicles. At this time slot cell 1 has occupancy and flow capacity of 32 and 48 respectively and doesn't apply any restriction to the upcoming flows from the cell 101 or 102. It should be further noted that both the cells have a penalty of 3 on the occupancies. As a result, the remaining 13.5 vehicles in the cell 102 can be moved to cell 101 as it wouldn't change the objective value. This is shown in Figure 2(b). However, it is not possible to find the same objective value if we would keep this 13.5 vehicles in the source cell R as the source cell has a higher penalty on occupancy. Furthermore, this 13.5 demands cannot be in the cell one as it would attain better solution than SO which is impossible.



(a) The optimal solution. (b) Constructed solution.

Figure 4.2. Figures show the optimal solution of the NHB SO-DTA framework at time slot $t = 18$ and one of the possible alternative solutions by interchanging the occupancies at the same penalty level. The numbers in the blue-boxes and ellipses are occupancies and flows, respectively.

4.3.2 Optimal Flows at the Penalty Levels

The optimal flow structure stating that the NHB SO-DTA formulation attains maximum possible flow at each of the penalty levels was discussed in Theorem 4.2.4. Here, we are going to present numerical optimal solution related to this structure. Like the previous section let's consider the 60s CTM network presented in Figure 1(a) with the same physical parameters and demand profile as before. The optimal solution at time-slot $t = 3$ is presented in Figure 4.3. In the figure, it is noticeable that the flows are always maximum as one of the inequality constraints (3.2)-(3.12) are tight. For instance, at the cells 101 and 102, the occupancy values are 36 and 40, respectively. However, already 30 vehicles are occupying cell 1 and it has total occupancy capacity of 96 (please see Table. 4.1). According to constraint (3.11), at this time-slot a maximum number of vehicles that can be taken from cells 101 and 102 is 30. In the Figure 4.3 we see that the optimal flows from cell 101 and 102 are 22.5 and 7.5, respectively. The combined flow from these two cells is 30 which is the maximum attainable flow at time-slot $t = 3$ due to the restriction on the flow by the queue spillback constraint in (3.11). However, for this particular case if we try to interchange the optimal flows within the same penalty level (e.g., cell 101 and

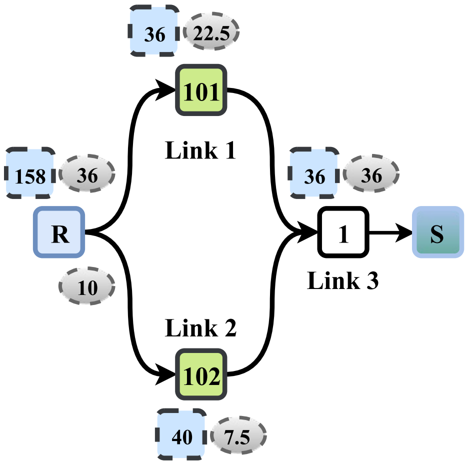


Figure 4.3. The optimal occupancies (in rectangles) and flows (in ellipsoids) at the time-slot $t = 3$. At the each of the penalty level (e.g., cells 101 and 102), the optimal flows are the maximum attainable flows at that time-slot.

102), then this would force other optimal flows and/or occupancies to change due to the CTM balance equations or constraints. In congested traffic conditions like over-saturated traffic condition or queue spill-back case, it is very difficult to interchange optimal flows and construct a new solution, though it might be straight forward in an under-saturated traffic scenario.

4.3.3 Delay Minimization

In the previous section, we have discussed that the NHB SO-DTA formulation maximizes flow at the each of the penalty levels. Furthermore, the aggregated occupancy at the individual penalty level has to be the same. As a matter of fact, the delay at the each of the penalty level is minimized as aggregated flow and occupancy is not changeable among different penalty levels during that time-slot. This structure was presented in Theorem 4.2.5. For numerical example let's consider the optimal solution presented in the Figure 4.3. In the figure, we see that at the penalty level of cell 101 and 102 the aggregated occupancy is 76 and the total outgoing flow from the level is 30, which is maximum. As a result, the delay/residual queue at $t = 3$ at that penalty level is 46. This is also noticeable that we cannot interchange the occupancy

Table 4.2. Intersection delay for different formulations.

CTM network	Formulation	Total intersection delay
20s CTM	HB SO-DTA	107.5
	NHB SO-DTA	1126
	NHB IDM	1126
60s CTM	HB SO-DTA	20.6
	NHB SO-DTA	916.5
	NHB IDM	916.5

among different levels (e.g., cell R and 101) because they have different penalty value on the objective function. Hence, the delay is also minimized at that penalty level and time-slot.

The numerical example of Theorem 4.2.6 is a straight forward extension of the above fact. If we consider the individual penalty level delay minimization property of the NHB SO-DTA formulation for all the time-slots and all the available penalty levels in the network, then one can easily deduce that the total system-wide delay is also minimized.

Now we discuss the Intersection Delay Minimization (IDM) structure presented in Lemma 4.2.7. To measure intersection delay for the formulations $P_{NHB\ SO\text{-}DTA}$ and $P2$, we have used both the 60s CTM and 20s CTM networks presented in Figure 4.1. The physical parameters are outlined in Table. 4.1. The demand profile used for the 60s CTM network is {90, 180, and 270} [vehicles] in the first consecutive time slots. Whereas the generated demand at the source cell for the 20s CTM network is {30, 30, 30, 60, 60, 60, 90, 90, and 90} [vehicles] during the first nine time slots. Three different formulations were used to solve the mentioned networks. In Table. 4.2 we see HB SO-DTA formulation attains lowest intersection delay. This is due to the fact that this formulation holds back most of the traffic in the source cell. On the other hand, both the NHB SO-DTA and NHB IDM attain same intersection delay. This result verifies the structure presented in Lemma 4.2.7.

4.4 The NHB DTA-SC Formulation With Active Control

In this section, firstly the Non-Holding-Back Dynamic Traffic Assignment with Signal Control (NHB DTA-SC) formulation is presented in which signal control is always active. Optimal solution of this formulation will be studied to analyze optimal signal control structures. The notations used in this section are defined below.

Decision variables, sets, and indexes

x_i^t	occupancy of cell i at time-slot t
y_{ij}^t	flow from cell i to j at time-slot t
g_i^t	green-split allocated to intersection cell i at time-slot t
$\mu_{ij}^{1t}, \mu_{ij}^{3t}, \pi_i^{2t}, \pi_i^{4t}, \pi_{(z,p)}^t$	are non-negative Lagrange multipliers
$\lambda_i^t, \theta_i^t, \sigma_{(z,p)}^t$	are Lagrange multipliers unrestricted in sign
$\mathcal{I}$	set of all the intersections where I is an intersection $\in \mathcal{I}$
$\mathcal{I}$	set of all the intersections I
$\mathcal{C}_I$	set of all the intersection cells
$\mathcal{C}_I^+$	subset of $\mathcal{C}_I$ which has green-time strictly greater than zero
$\mathcal{C}_{IM}$	set of intersection merge cells
$\mathcal{C}_{IM}^+$	set of intersection merge cells in spillback condition
P	set of all the phases
P^+	the phases $\in P$ that have green-time greater than zero
I^+	an intersection $\in \mathcal{I}$ in a specific traffic condition
z	intersection index
p	a phase in P
i, j, k	cell index

The major focus of this section is to derive solution structures of an NHB DTA-SC formulation. We propose a linear objective function that not only maximizes throughput of a network, but also makes the study of traffic signal control relevant. The complete NHB DTA-SC formulation is presented below.

$P_{NHB\ DTA\text{-}SC}$:

$$\text{maximize} \qquad \mathcal{O}_1 = \sum_{\forall t \in \mathcal{T}} \sum_{\forall ij \in \mathcal{E}_S} y_{ij}^t - \sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C} \setminus \mathcal{C}_S} \Psi_i \cdot x_i^t - \sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C}_I} g_i^t \tag{4.19}$$

subject to

$$\sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C} \setminus \mathcal{C}_S} x_i^t = \mathcal{O}_Z^{\star}, \tag{4.20}$$

$$x_i^t = x_i^{t-1} + d_i^{t-1} - y_{ij}^{t-1}, \; \forall i \in \mathcal{C}_R, \; j \in \Gamma(i), \tag{4.21}$$

$$x_i^t = x_i^{t-1} + y_{ki}^{t-1} - y_{ij}^{t-1}, \; \forall ij \in \mathcal{E}_O, \; j \in \Gamma(i), \tag{4.22}$$

$$x_i^t = x_i^{t-1} + y_{ki}^{t-1} - \sum_{j \in \Gamma(i)} y_{ij}^{t-1}, \; \forall i \in \mathcal{C}_D, \; k \in \Gamma^{-1}(i), \tag{4.23}$$

$$x_i^t = x_i^{t-1} + \sum_{k \in \Gamma^{-1}(i)} y_{ki}^{t-1} - y_{ij}^{t-1}, \; \forall i \in \mathcal{C}_M \cup \mathcal{C}_{IM}, \tag{4.24}$$

$$y_{ij}^t - x_i^t \leq 0, \; y_{ij}^t \leq Q_i^t, \; \forall ij \in \mathcal{E}_{R \cup O}, \; j \in \Gamma(i), \tag{4.25}$$

$$y_{ij}^t \leq Q_j^t, \; y_{ij}^t \leq \delta_j^t (N_j^t - x_j^t), \; \forall ij \in \mathcal{E}_{R \cup O \cup D}, \; j \in \Gamma(i), \tag{4.26}$$

$$\sum_{j \in \Gamma(i)} y_{ij}^t - x_i^t \leq 0, \; \sum_{j \in \Gamma(i)} y_{ij}^t \leq Q_i^t, \; \forall ij \in \mathcal{E}_D, \; j \in \Gamma(i), \tag{4.27}$$

$$y_{ki}^t \leq Q_k^t, \; \forall ki \in \mathcal{E}_{M \cup S}, \; k \in \Gamma^{-1}(i), \tag{4.28}$$

$$y_{ki}^t \leq x_k^t, \; \forall ki \in \mathcal{E}_{M \cup S \cup IM}, \; k \in \Gamma^{-1}(i), \tag{4.29}$$

$$\sum_{k \in \Gamma^{-1}(i)} y_{ki}^t \leq \delta_i^t (N_i^t - x_i^t), \; \forall ki \in \mathcal{E}_M \cup \mathcal{E}_{IM}, \; k \in \Gamma^{-1}(i), \tag{4.30}$$

$$\sum_{k \in \Gamma^{-1}(i)} y_{ki}^t \leq Q_i^t, \; \forall ki \in \mathcal{E}_M \cup \mathcal{E}_{IM}, \; k \in \Gamma^{-1}(i), \tag{4.31}$$

$$y_{ij}^t - x_i^t \leq 0, \; y_{ij}^t \leq g_i^t Q_i^t, \; \forall i \in \mathcal{C}_I, \; \forall j \in \Gamma(i), \tag{4.32}$$

$$g_{(z,p)}^t = g_{(z,p+4)}^t, \; \forall (z,p) \in \mathcal{C}_I, \; \forall z \in \mathcal{I}, \; \forall p \in P, \tag{4.33}$$

$$g_{(z,2p)}^t = g_{(z,p+8)}^t, \; \forall (z,p) \in \mathcal{C}_I, \; \forall z \in \mathcal{I}, \; \forall p \in P, \tag{4.34}$$

$$\sum_{p \in P} g_{(z,p)}^t \leqslant 1, \; \forall (z,p) \in \mathcal{C}_I, \; \forall z \in \mathcal{I}, \; \forall p \in P, \tag{4.35}$$

$$\forall t \in \mathcal{T}.$$

The implications of the objective and constraints are discussed next. In Section 3.2.2.1, we have provided an algorithm to generate the penalty terms (Ψ_is) followed by an example of it's application in a cyclic network.

The linear objective function in (4.19) not only can successfully resolve the HB problem but also forces to attain effective green-splits regardless of the traffic conditions due to the inclusion of g_i^t term. Effective green-splits are the exact proportions required to serve the optimal flows. Mathematically this can be written as,

$y_{ij}^t = G_i^{t^{eff}} Q_i^t, \ \forall i \in \mathcal{C}_I, \ \forall j \in \Gamma(i), \ \forall t \in \mathcal{T}$. The above formulation always attains $g_i^{\star t} = G_i^{t^{eff}}$, where $g_i^{\star t}$ is the optimal green-split. Note that if we would not add the g_i^t term in the objective function, in under-saturated and spill-back conditions the signal control constraints would become inactive. More details about this fact will be provided in the next section. Also note that considering minimum green-time g_{min} which is the lower bound of the green-splits and fixed for all the phases (i.e., at every cycle all the phases must get g_{min} proportion of green-time) would not always assure active control. For instance, in an under-saturated condition let us assume that $g_{min} = 0.1$ and it is required to allocated $g_i^t = 0.5$. In this case, if we allocated $g_i^{\star t} = 1$ then the signal control does not play a role in the optimal solution and becomes inactive.

The first term in the objective represents the total flow into the sink cells, and hence, its maximization is equivalent to the maximization of the total network-wise throughput. The second term represents a penalty where an artificial penalty Ψ_i is multiplied with the occupancy at each cell i. The idea is that the vehicle holding-back (HB) problem can be eliminated by adding a cell-dependent penalty at each cell. The penalty Ψ_i is predefined. The highest penalty is associated with the source cells and the penalty decreases as one moves closer towards the destination. If cell i is upstream to cell j, then $\Psi_i > \Psi_j$.

The third term in the objective function forces the signal control constraints to become always active. We have found that the most of the formulations in the literature would make the signal control inactive in an under-saturated or spillback traffic condition. This is because in the mentioned traffic conditions it is not required to allocate the total green-time. As a matter of fact, in those formulations any green-time that does not restrict the optimal flow is acceptable. This makes the attained green-proportions very arbitrary and does not represent effective green-time that is required to serve the optimal flow.

To ensure System Optimal (SO) property of the solution, we add the constraint

(4.20) in the formulation, where $\mathcal{O}^{\star}_{Z}$ represents the optimal objective value attained by solving the following SO-DTA problem (Ziliaskopoulos 2000).

$$\mathcal{O}_Z = \sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C} \setminus \mathcal{C}_S} x_i^t, \tag{4.36}$$

Subject to the constraints (4.21)-(4.35).

The constraints (4.21)-(4.24) represent the linear conservation equations or the flow-occupancy balance relations for different types of cells, and (4.25)-(4.31) are linear relaxations to the non-linear CTM flow propagation rules introduced by Daganzo (1994).

Constraints (4.32)-(4.35) represent traffic signal control model proposed by Ukkusuri, Ramadurai, and Patil (2010), where g_i^t denotes the green split allocated to the traffic from an intersection cell i, $i \in \mathcal{C}_I$, at time slot t. The green split g_i^t essentially restricts the maximum flow capacity Q_i^t for the traffic from the intersection cell i, $i \in \mathcal{C}_I$, and leads to the constraint (4.32). Constraint (4.33) defines signal phasing convention prescribed by the national electrical manufacturers association (NEMA). The NEMA convention mandates that the phase pairs of (1,5), (2,6), (3,7), and (4,8) are allowed together. Also, one of the combinations (2,9), (4,10), (6,11), or (8,12) is allowed at a time if right turn is allowed and no conflicts occur which is modeled by constraint (4.34). Note that, in these constraints the index z is an intersection index and p is the phase index. Both of these indices combinedly form an intersection cell index. For example, if the $z = 1$ and $p = 1$ then the intersection cell index $i = (1, 1) = 101$. Similarly, for the other phases the intersection cell index would be 102, 103, 104, respectively. Finally, one must ensure that all the sum of the green fractions cannot exceed 1, which is ensured by the constraint (4.35).

It is important to mention that the major contributions of this chapter are not the proposed NHB DTA-SC formulation in $P_{NHB\ DTA\text{-}SC}$ but the decomposition of $P_{NHB\ DTA\text{-}SC}$ into separate sub-problems and then study the locally optimal signal control structures in the formulation under different traffic conditions. In the following section, we present the decomposed sub-problems and the structures of the optimal traffic signal settings in over/under-saturated and spillback traffic conditions.

4.5 Decomposition of the NHB DTA-SC Model

In this section, primarily we decompose the NHB DTA-SC formulation presented in Section 4.4 into three separate subproblems. The sub-problems are occupancy minimization (OM), flow maximization (FM), and signal control (SC). Since the major focus of this study is to analyze signal control and its optimal structure, we consider these sub-problems and the corresponding decision variables only at the intersection. Nevertheless, these sub-problems can be easily extended to apply to the whole network. This can be done simply by taking the summations over the whole network rather than considering intersection cells only. Then we further analyze the SC sub-problem. We derive Karush-Kuhn-Tucker (KKT) optimality conditions in the signal control problem and present locally optimal structures of the traffic signal control in over-saturated, under-saturated, and spillback traffic scenarios.

4.5.1 Decomposition of the Problem Into Subproblems

To analyze the optimal structure of the traffic signal control, let's consider the Lagrangian dual function presented in Equation (4.37). Since our primary concern is to report the optimal structure of the traffic signal control, we analyze the Lagrangian at the intersection cells only. By using the formulation in $P_{NHB\ DTA\text{-}SC}$ one can write a Lagrange function $\mathcal{L}_i$ at an intersection cell i at time t as follows

$$\begin{aligned}
\mathcal{L}_i = & -\Psi_i \cdot x_i^t - \Psi_i \cdot g_i^t && (4.37)\\
& + (x_i^t - x_i^{t-1} - y_{ki}^{t-1} + y_{ij}^{t-1})\lambda_i^t && (4.38)\\
& + (\mathcal{O}_Z^\star - x_i^t)\theta_i^t && (4.39)\\
& + (x_i^t - y_{ij}^t)\mu_{ij}^{1t} && (4.40)
\end{aligned}$$

$$+ (g_i^t Q_i^t - y_{ij}^t)\mu_{ij}^{3t} \tag{4.41}$$

$$+ (Q_j^t - \sum_{i \in \Gamma^{-1}(j)} y_{ij}^t)\pi_i^{2t} \tag{4.42}$$

$$+ \{\delta_j^t(N_j^t - x_j^t) - \sum_{i \in \Gamma^{-1}(j)} y_{ij}^t\}\pi_i^{4t} \tag{4.43}$$

$$+ (1 - \sum_{p=1}^{4} g_{z,p}^t)\pi_{(z,p)}^t \tag{4.44}$$

$$+ (g_{(z,p+4)}^t - g_{(z,p)}^t)\sigma_{(z,p)}^t, \tag{4.45}$$

$k \in \Gamma^{-1}(i), j \in \Gamma(i), \forall (z,p) \in i,\ \forall i \in \mathcal{C}_I,\ \forall j \in \mathcal{C}_{IM}, x_i^t, y_{ij}^t, g_i^t, \mu_{ij}^{1t}, \mu_{ij}^{3t}, \pi_i^{2t}, \pi_i^{4t}, \pi_{(z,p)}^t \geq 0, \lambda_i^t, \theta_i^t, \sigma_{(z,p)}^t$ unrestricted in sign.

In the above dual function we ignore the coordination between through and right movements defined by (4.34). This is to simplify the Lagrangian dual function and obtain insights. The Lagrange multipliers $\lambda_i^t, \theta_i^t, \mu_{ij}^{1t}, \mu_{ij}^{3t}, \pi_i^{2t}, \pi_i^{4t}, \pi_{(z,p)}^t, \sigma_{(z,p)}^t$ solve the problem (Shakkottai, Srikant et al. 2008)

$$\min_{(\lambda_i^t, \theta_i^t, \mu_{ij}^{1t}, \mu_{ij}^{3t}, \pi_i^{2t}, \pi_i^{4t}, \pi_{(z,p)}^t, \sigma_{(z,p)}^t)} \mathcal{D}(\lambda_i^t, \theta_i^t, \mu_{ij}^{1t}, \mu_{ij}^{3t}, \pi_i^{2t}, \pi_i^{4t}, \pi_{(z,p)}^t, \sigma_{(z,p)}^t),$$

where

$$\mathcal{D}(\lambda_i^t, \theta_i^t, \mu_{ij}^{1t}, \mu_{ij}^{3t}, \pi_i^{2t}, \pi_i^{4t}, \pi_{(z,p)}^t, \sigma_{(z,p)}^t) = \max_{x_i^t, y_{ij}^t, g_i^t} \mathcal{L}_i.$$

It can be observed that the Lagrange function (4.37) is linear in the primal variables, i.e., there are no product terms of primal variables involved. As a result, applying *dual decomposition*, we can decompose the problem into different sub-problems. Resources such as green time at the intersection can be allocated via so-called *pricing*, and the prices at the different layer of decomposition connect the sub-problems (Palomar and Chiang 2006). Specifically, we separate the Lagrangian function into three different local sub-problems that are occupancy minimization (OM), flow maximization (FM), and utilization of signal control (SC). The sub-problems are linked through the so-called *shadow prices* that are essentially the values of the Lagrange multipliers at the optimal point. The decomposed OM and FM sub-problems are presented below.

$$\mathcal{D}_{OM}(\lambda_i^t, \theta_i^t, \mu_{ij}^{1t}, \pi_k^{4t}) := \max_{x_i^t \geq 0} x_i^t \lambda_i^t - \Psi_i \cdot x_i^t - x_i^{t-1}\lambda_i^{t-1} - x_i^t\theta_i^t + x_i^t\mu_{ij}^{1t} - x_i^t\pi_k^{4t}. \quad (4.46)$$

$$\mathcal{D}_{FM}(\lambda_i^t, \lambda_j^t, \mu_{ij}^{1t}, \mu_{ij}^{3t}, \pi_i^{2t}, \pi_i^{4t}) := \max_{y_{ij}^t \geq 0} y_{ij}^t(-\lambda_j^t - \mu_{ij}^{1t} - \mu_{ij}^{3t} - \pi_i^{2t} - \pi_i^{4t}) + y_{ij}^{t-1}\lambda_i^{t-1}. \quad (4.47)$$

The above two sub-problems are connected via the marginal cost λ_i^{t-1} and the benefit μ_{ij}^{1t} of moving one unit of occupancy from the cell i to cell j. In the subproblem (4.46) all other terms are negative except the terms related to the Lagrange multipliers μ_{ij}^{1t} and λ_i^t. The dual variable λ_i^t is associated with an equality constraint and, therefore, is unrestricted in sign and the term related to this price is $-\infty$ when $\lambda_i^t = -\infty$ or 0 when $\lambda_i^t = 0$. In both the cases the marginal cost is equal at time $t-1$. To maximize the sub-problem $\mathcal{D}_{OM}$ one needs to maximize the term $x_i^t\mu_{ij}^{1t}$. However, increasing x_i^t would result in lower value of $\mathcal{D}_{OM}$ as $\Psi_i > 0$, which forces the μ_{ij}^{1t} to be nonzero and the flow to be higher. This corresponding relation between the two sub-problems implies occupancy to be minimum at the higher penalty levels and flow to be maximum attainable flow at the cells. Due to this connect between OM and FM, the formulation $P_{NHB\ DTA\text{-}SC}$ attains NHB solution. However, a situation may arise when $\mu_{ij}^{1t} = 0$ and the corresponding constraint will be non-binding (i.e., $x_i^t > y_{ij}^t$). Then the optimal flow would be determined by other constraints related to the flow and the marginal cost. The decomposition related to the traffic signal control variable can be written as,

$$\mathcal{D}_{SC}(\mu_{ij}^{3t}, \pi_{(z,p)}^t, \sigma_{(z,p)}^t) := \max_{1 \geq g_i^t \geq 0} -\Psi_i g_i^t + g_i^t Q_i^t \mu_{ij}^{3t} - g_i^t \pi_{(z,p)}^t - g_i^t \sigma_{(z,p)}^t. \quad (4.48)$$

In the above decomposed signal control problem, $\mathcal{D}_{SC}$ is to be maximized for the minimum values of the associated Lagrange multipliers. The term $\Psi_i g_i^t$ penalizes the objective function of the primal DTA-SC problem in $P_{NHB\ DTA\text{-}SC}$ and ensures that the constraint in Equation (4.32) is always active regardless of the traffic conditions. The second term in Equation (4.48) couples the green-split allocation problem to the

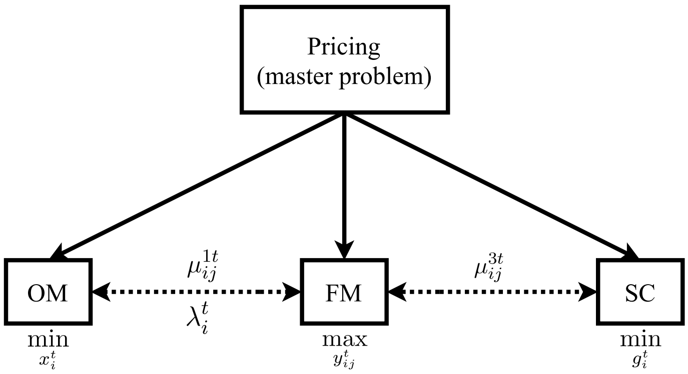


Figure 4.4. The master problem in the figure determines the prices λ_i^t, λ_i^{t-1}, θ_i^t, μ_{ij}^{1t}, π_k^{4t} for the sub-problem OM; the prices λ_j^t, μ_{ij}^{1t}, μ_{ij}^{3t}, π_i^{2t}, π_i^{4t}, λ_i^{t-1} for the sub-problem FM; and the prices μ_{ij}^{3t}, $\pi_{(z,p)}^t$, $\sigma_{(z,p)}^t$ for the SC sub-problem. The OM, FM, and SC sub-problems are connected via the so-called shadow prices that are the values of the Lagrange multipliers at the optimal point and each shadow price can be interpreted as the marginal cost for violating a specific constraint by one unit.

flow maximization problem in (4.47) through μ_{ij}^{3t}. This price is determined by $\mathcal{D}_{OM}$ as this sub-problem decides how much resources (i.e. green-time) should be allocated to serve the optimal flow. Figure 4.4 shows the connections among the subproblems via the corresponding Lagrange multipliers. The Lagrange multiplier $\pi_{(z,p)}^t$ is the price of the allocated green-split from the total available green proportion. $\sigma_{(z,p)}^t$ is the phase coordination cost to allocate equal green-splits to the traffic movements belong to the same phase and to avoid any intra-phase conflicts. Now we need to find the stationary points in Equation (4.48). In this sub problem $\sigma_{(z,p)}^t$ is unrestricted in sign and notice that $g_i^t \sigma_{(z,p)}^t = +\infty$ unless $\sigma_{(z,p)}^t = 0$. This reduces the signal control problem to

$$\mathcal{D}_{SC}(\mu_{ij}^{3t}, \pi_{(z,p)}^t, \sigma_{(z,p)}^t) := \max_{1 \geq g_i^t \geq 0} -\Psi_i g_i^t + g_i^t Q_i^t \mu_{ij}^{3t} - g_i^t \pi_{(z,p)}^t. \tag{4.49}$$

4.5.2 Karush-Kuhn-Tucker (KKT) Optimality Conditions

The NHB DTA-SC framework in $P_{NHB\ DTA\text{-}SC}$ is linear, primal feasible, and convex. Furthermore, the weak form of Slater's condition holds in different traffic conditions. As a result, there is zero duality gap (Boyd and Vandenberghe 2004) between the primal and dual problem of the NHB DTA-SC framework.

The signal control models added in the formulation are convex (i.e., linear) and differentiable. Due to the NHB nature of the formulation at least one of the flow constraints at an intersection cell must be satisfied by an equality constraint and the corresponding Lagrange multiplier is positive. As a consequence, few affine constraints always exist at an intersection. Hence, any points that satisfy the KKT conditions of the signal control sub-problem (4.49) are primal and dual optimal (Boyd and Vandenberghe 2004).

It is to be noted that if we use the KKT optimality conditions and solve the signal control sub-problem, this is equivalent to satisfying the complementary slackness conditions and finding the stationary points of the Lagrangian (Chiang et al. 2007) in Equation (4.49). When there is non-zero flow and occupancy at that intersection cell at the time-slot t, the KKT optimality conditions related to the sub-problem $\mathcal{D}_{SC}(\mu_{ij}^{3t}, \pi_{(z,p)}^{t}, \sigma_{(z,p)}^{t})$ can be written as,

$$\frac{\partial \mathcal{L}_i}{\partial g_i^{\star t}} = -1 + Q_i^t \mu_{ij}^{\star 3t} - \pi_{(z,p)}^{\star t} = 0 \Rightarrow \pi_{(z,p)}^{\star t} = Q_i^t \mu_{ij}^{\star 3t} - 1, \tag{4.50}$$

$$(g_i^{\star t} \cdot Q_i^t - y_{ij}^{\star t})\mu_{ij}^{\star 3t} = 0, \tag{4.51}$$

$$(1 - \sum_{p=1}^{4} g_{(z,p)}^{\star t})\pi_{(z,p)}^{\star t} = 0, \tag{4.52}$$

where the primal variables $g_i^t, y_{ij}^t \geq 0$ and the dual variables $\mu_{ij}^{3t}, \pi_{(z,p)}^t \geq 0$. The '$\star$' superscript on a primal or dual variable indicates optimal value of the corresponding variable. Based on the different values of these Lagrange multipliers the optimal traffic signal setting for different traffic conditions can be determined. Two important properties of the above KKT conditions are discussed below.

Remark 2. *It is evident in the above KKT conditions that the SC sub-problem leads to locally optimal decisions of green-splits. Applying KKT conditions, it can be similarly shown that the OM and FM sub-problems also lead to corresponding locally optimal decisions of occupancies and inter-cell flows. It is both important and interesting to observe that such locally optimum decisions ultimately lead to globally optimal solutions such as NHB DTA-SC.* $\odot$

Remark 3. *In the KKT optimality condition* (4.50) *notice that if we would not penalize objective function for the allocated green-split then* 1 *would not be present in the condition. Consequently, the signal control constraint related to complementary slackness condition* (4.51) *would become inactive in an under-saturated or spill-back traffic condition. This is because, in the mentioned traffic scenarios it is not required to allocate the total green-time. As a result,* $\pi^{\star t}_{(z,p)} = 0$*, which implies* $g^{\star t}_i \cdot Q^t_i > y^{\star t}_{ij}$ *in* (4.51) *and inactive. To explain this fact better, let us take the Total System-wide Travel Time (TSTT) objective function presented by Ziliaskopoulos (2000),*

$$\mathcal{O}_Z = \sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C} \setminus \mathcal{C}_S} x^t_i, \; \forall t \in \mathcal{T}. \tag{4.53}$$

While using the above objective function one would get the first order differential KKT condition

$$\frac{\partial \mathcal{L}_i}{\partial g^{\star t}_i} = Q^t_i \mu^{\star 3t}_{ij} - \pi^{\star t}_{(z,p)} = 0 \Rightarrow \pi^{\star t}_{(z,p)} = Q^t_i \mu^{\star 3t}_{ij}, \; \forall i \in \mathcal{C}_I, \; \forall t \in \mathcal{T}. \tag{4.54}$$

In an under-saturated or spill-back traffic condition it is not required to allocate the total green-time and

$$\sum_{p=1}^{4} g^{\star t}_{(z,p)} < 1, \; (z,p) \in \mathcal{C}_I, \; z \in \mathcal{I}, \; \forall p \in P, \; t \in \mathcal{T}. \tag{4.55}$$

By combining above with the corresponding complementary slackness condition one can deduce $\pi^{\star t}_{(z,p)} = 0$*. By substituting this value in the Equation* (4.54) *one gets* $\mu^{\star 3t}_{ij} = 0$*. As a result, from the complementary slackness condition (like one in* (4.51)*) we can infer*

$$y^{\star t}_{ij} < g^{\star t}_i \cdot Q^t_i, \; i \in \mathcal{C}_I, \; t \in \mathcal{T}. \tag{4.56}$$

Consequently, both of the signal control constraints become inactive. $\odot$

In the following section we are going to discuss three different optimal control structures in different traffic states.

4.5.3 Structures of the Optimal Traffic Signal Control

The major contributions of this chapter are optimal traffic signal control structures in the NHB DTA-SC formulation presented in $P_{NHB\ DTA\text{-}SC}$. In the following, we state and prove different structures of the optimal signal control settings in over-saturated, under-saturated, and spill-back traffic conditions. We apply KKT optimality conditions of the decomposed SC problem presented above to derive these optimal control structures.

The over-saturated condition is one of the most important traffic conditions under which the overall network-wide throughput critically depends on the green splits determined by signal control. For this reason we will study below the structure of NHB DTA-SC solution in the over-saturated traffic condition where we assume that the intersections are the bottlenecks and the links in the network have sufficient flow and occupancy capacity to accommodate the flows dictated by that solution. In the following, we first define the over-saturated traffic condition and then state and prove the main structural result for such traffic conditions.

Definition 4.5.1 (Over-saturated Intersection)**.** The over-saturated condition implies that the occupancy at each intersection cell exceeds its flow capacity such that a part of the cell occupancy cannot be cleared and the formation of queue occurs, and the following is true in such conditions

$$x_i^t > y_{ij}^t,\ \forall i \in \mathcal{C}_I,\ \forall j \in \mathcal{C}_{IM},\ \forall t \in \mathcal{T}, \tag{4.57}$$

$$\max_i Q_i^t \leq Q_j^t,\ \forall i \in \mathcal{C}_I,\ \forall j \in \mathcal{C}_{IM},\ \forall t \in \mathcal{T}, \tag{4.58}$$

$$\sum_{\forall i \in \Gamma^{-1}(j)} y_{ij}^t < \delta_j^t (N_j^t - x_j^t),\ \forall i \in \mathcal{C}_I,\ \forall j \in \mathcal{C}_{IM},\ \forall t \in \mathcal{T}. \tag{4.59}$$

As a result, to maximize flow through the intersection (or minimize occupancy) the total green-time must be allocated at optimal point, which is

$$\sum_{p=1}^{4} g_{(z,p)}^{\star t} = 1. \tag{4.60}$$

□

Proposition 4.5.2. *In an over-saturated intersection, if the only bottleneck in the network is the intersection, then the structure of the optimal green splits g_i^t in an NHB DTA-SC formulation have the structure of the Webster's control method (Webster 1958).*

Proof. In an over-saturated traffic condition, the flow at the intersection is nonzero and there exists at least an $i \in \mathcal{C}_I$ such that $g_i^{\star t} > 0$. We consider the set $\mathcal{C}_I^+(\subset \mathcal{C}_I)$, such that $g_i^{\star t} > 0$ for all $i \in \mathcal{C}_I^+$, i.e., $\mathcal{C}_I^+$ is a set of intersection cells in which all the cells have green-splits greater than zero. Then, $\forall i \in \mathcal{C}_I^+$, we can apply $g_i^{\star t} > 0$, and the green-split $g_i^{\star t}$ must satisfy the KKT condition

$$\pi_{(z,p)}^{\star t} = Q_i^t \mu_{ij}^{\star 3t} - 1,\ \forall i \in \mathcal{C}_I^+,\ \forall ij \in \mathcal{E}_{IM},\ \forall i \in (z,p)\,,\ \forall t \in \mathcal{T}. \tag{4.61}$$

As discussed before, z is the intersection index and p is the phase index that combinedly (i.e., (z,p)) represent an intersection cell index i. The Definition 4.5.1 implies that the constraint related to complementary slackness condition (4.52) is tight and $\pi_{(z,p)}^{\star t} > 0$. By combining this value of the Lagrange multiplier with (4.61) we get

$$Q_i^t \mu_{ij}^{\star 3t} - 1 > 0. \tag{4.62}$$

As both the Q_i^t and 1 are positive, one can deduce from (4.62) that

$$\mu_{ij}^{\star 3t} > 0. \tag{4.63}$$

By combining above with the complementary slackness condition (4.51) we get

$$y_{ij}^{\star t} = g_i^{\star t} Q_i^t \Rightarrow g_i^{\star t} = \frac{y_{ij}^{\star t}}{Q_i^t},\ \forall i \in \mathcal{C}_I^+,\ \forall ij \in \mathcal{E}_{IM},\ \forall i \in (z,p)\,,\ \forall t \in \mathcal{T}. \tag{4.64}$$

Rewriting (4.60) as

$$\sum_{p \in P^+} g_{(z,p)}^{\star t} = 1,\ \forall i \in \mathcal{C}_I^+,\ \forall ij \in \mathcal{E}_{IM},\ \forall i \in (z,p)\,,\ \forall t \in \mathcal{T}, \tag{4.65}$$

where P^+ is a subset of the phases P with strictly positive green times. Combining above with (4.64), we can write

$$\sum_{p \in P^+} g^{\star t}_{(z,p)} = \sum_{p \in P^+} \frac{y^{\star t}_{(z,p)j}}{Q^t_{z,p}} = 1, \ \forall (z,p) \in \mathcal{C}^+_I, \ p \in P, \ j \in \Gamma_{(z,p)}, \ \forall t \in \mathcal{T}. \tag{4.66}$$

By substituting (4.66) in (4.64), we get

$$g^{\star t}_{(z,p)} = \frac{\frac{y^{\star t}_{(z,p)j}}{Q^t_{z,p}}}{\sum_{p \in P^+} \frac{y^{\star t}_{(z,p)j}}{Q^t_{z,p}}}, \ \forall (z,p) \in \mathcal{C}^+_I, \ p \in P^+, \ j \in \Gamma_{(z,p)}, \ \forall t \in \mathcal{T}. \tag{4.67}$$

The above equation shows that the optimal control at the NHB DTA-SC framework follows Webster's control $\forall i \in \mathcal{C}^+_I$ in over-saturated condition. For $\forall i \in \mathcal{C}_I \setminus \mathcal{C}^+_I$, we have $g^{\star t}_i = y^{\star t}_{ij} = 0$, and Webster's control structure trivially holds. It is to be noted that Webster (1958) signal control locally maximizes the throughput of the intersection. Hence, the optimal control in the NHB DTA-SC formulation is also throughput optimal under the assumptions stated in the Definition 4.5.1. □

Smith (1980) showed that in an oversaturated condition P_0 control policy is optimal when a UE route-choice behavior is considered. The optimal control structure we have derived above is from a SO setting and should not be confused with Smith's result. Similar to our result, max-pressure control policy (Varaiya 2013) minimizes intersection delay. It allocates green-splits based on pressure at upstream intersection cells when the pressure is linearly dependent on queue length. In the proposed NHB DTA-SC framework, occupancy is interchangeable within the same penalty level and it maximizes flow. As a result, according to Lemma 4.2.7 it also minimizes intersection delay.

Over-saturated condition occurs when heavy demands arrive at an intersection than the total serving capacity of it. This state of traffic may appear during the morning and afternoon peak hours in a working day. Unlike over-saturated condition, under-saturated condition occurs when there is very light traffic arriving at the intersection and it is not required to utilize the total capacity of the intersection. This traffic state normally arises during the nights of the working days. In the following, we analyze the structure of an optimal signal control setting under this light demand traffic condition. We define this traffic condition at the intersection in the following way.

Definition 4.5.3 (Under-saturated Intersection)**.** An under-saturated traffic condition at an intersection can be defined using the following conditions in the NHB DTA-SC framework,

$$x_i^t = y_{ij}^t < Q_i^t,\ \forall i \in \mathcal{C}_I,\ \forall j \in \mathcal{C}_{IM},\ \forall t \in \mathcal{T}, \tag{4.68}$$

$$\sum_{\forall i \in \Gamma^{-1}(j)} y_{ij}^t < \delta_j^t (N_j^t - x_j^t),\ \forall i \in \mathcal{C}_I,\ \forall j \in \mathcal{C}_{IM},\ \forall t \in \mathcal{T}, \tag{4.69}$$

$$\sum_{\forall i \in \Gamma^{-1}(j)} y_{ij}^t < Q_j^t,\ \forall i \in \mathcal{C}_I,\ \forall j \in \mathcal{C}_{IM},\ \forall t \in \mathcal{T}, \tag{4.70}$$

$$\sum_{p \in P} g_{(z,p)}^{\star t} < 1,\ \forall (z,p) \in \mathcal{C}_I,\ \forall z \in \mathcal{I},\ \forall p \in P,\ \forall t \in \mathcal{T}. \tag{4.71}$$

△

The definition above implies that the occupancy at an intersection cell at the time-slot t can be served within that interval and there is no spill-back of traffic. Furthermore, at an under-saturated intersection it is not required to utilize the total green-time of the intersection and we assume the capacity constraint of the subsequent intersection merge cell does not impose any restriction on the flows. As a result, we add Equation (4.70) and (4.71) in the definition of this traffic condition.

Proposition 4.5.4. *At an under-saturated intersection I^+, when $\nexists j | \sum_{\forall i \in \Gamma^{-1}(j)} y_{ij}^t = \delta_j^t (N_j^t - x_j^t)$, $\sum_{\forall i \in \Gamma^{-1}(j)} y_{ij}^t = Q_j^t, \forall i \in \mathcal{C}_I^+, \forall j \in \mathcal{C}_{IM}^+, \forall ij \in I^+$ then the optimal green splits $g_i^{\star t}$ in the NHB DTA-SC ($P_{NHB\ DTA\text{-}SC}$) formulation are determined by the available occupancies $x_i^{\star t}$ at the intersection cells and $x_i^{\star t} = y_{ij}^{\star t} = g_i^{\star t} Q_i^t, \forall i \in \mathcal{C}_I^+$ holds.*

Proof. Let us recall the KKT condition (4.50) related to signal control variable which is

$$\pi^{\star t}_{(z,p)} = Q^t_i \mu^{\star 3t}_{ij} - 1,\ \forall i \in \mathcal{C}^+_I,\ \forall ij \in \mathcal{E}_{IM},\ \forall i \in (z,p)\,,\ \forall t \in \mathcal{T}. \tag{4.72}$$

In the above $\mathcal{C}^+_I \in I^+$ which indicates set of all the intersection cells belong to an under-saturated intersection I^+.

By combining the condition (4.71) in the Definition 4.5.3 and the complementary slackness condition (4.52), we get constraint (4.52) is not tight in an under-saturated condition and $\pi^{\star t}_{(z,p)} = 0$. By substituting this value of the Lagrange multiplier to (4.72), we get

$$Q^t_i \mu^{\star 3t}_{ij} = 1 \Rightarrow \mu^{\star 3t}_{ij} = \frac{1}{Q^t_i}. \tag{4.73}$$

In (4.73), both the 1 and Q^t_i are positive quantities which implies $\mu^{\star 3t}_{ij} > 0$. As a result, from the corresponding complementary slackness condition in (4.51) one can deduce

$$y^{\star t}_{ij} = g^{\star t}_i Q^t_i. \tag{4.74}$$

By combining (4.74) with the under-saturated condition in Equation (4.68) we get

$$x^{\star t}_i = y^{\star t}_{ij} = g^{\star t}_i Q^t_i. \tag{4.75}$$

This proves proportionality in between the optimal occupancy and green-time in an under-saturated condition and the Proposition 4.5.4 holds. □

From the under/over-saturated signal control structures presented above it is clear that the signal control constraint is always active in $P_{NHB\ DTA\text{-}SC}$ regardless of the traffic conditions. These structures couple signal control with the locally available demand. As a result, if the traffic arrival rate data (e.g., from camera, inductor loop) is available and the intersection capacity is known then based on the traffic conditions these structures can be plugged in to program a traffic signal system at an intersection.

Another signal control structure we are going to discuss here is when queue spills back from the downstream links. Whether a spillback condition will occur not only depends on the number of vehicles arriving at the intersection but also capacities at the downstream links. If we consider an isolated intersection then this scenario might arise due to the bottlenecks at the downstream links given that the intersection is in congested traffic state. We define the queue spillback condition in terms of the NHB DTA-SC constraints which is given below.

Definition 4.5.5 (Queue Spillback Condition)**.** This scenario can be mathematically defined using following equations:

$$x_i^t > y_{ij}^t,\ \forall i \in \mathcal{C}_I,\ \forall j \in \mathcal{C}_{IM},\ \forall t \in \mathcal{T}, \tag{4.76}$$

$$\sum_{\forall i \in \Gamma^{-1}(j)} y_{ij}^t = \delta_j^t(N_j^t - x_j^t) < Q_j^t,\ x_j^t > N_j^t - \delta_j^t N_j^t,\ \forall i \in \mathcal{C}_I,\ \forall j \in \mathcal{C}_{IM},\ \forall t \in \mathcal{T}, \tag{4.77}$$

$$\sum_{p \in P} g_{(z,p)}^{\star t} < 1,\ \forall (z,p) \in \mathcal{C}_I,\ \forall z \in \mathcal{I},\ \forall p \in P,\ \forall t \in \mathcal{T}. \tag{4.78}$$

△

In the above definition, (4.76) states that there is growing queue at the intersection cell i at time-slot t. The spillback condition is modeled by (4.77) as this equation implies how much flow can go out of the intersection cells are determined by the remainder occupancy capacities at the subsequent merge cells. As a result, similar to the under-saturated traffic condition, it is not required to allocate the total green-time of the intersection and the condition (4.78) is trivial. In the following we propose a structure of the optimal solution which relates optimal control with the discussed spillback traffic state.

Proposition 4.5.6. *At an intersection I^+ when queue spills back from the downstream link and $\exists j | \sum_{\forall i \in \Gamma^{-1}(j)} y_{ij}^t = \delta_j^t(N_j^t - x_j^t) < Q_j^t,\ x_j^t > Q_j^t,\ \forall i \in \mathcal{C}_I^+, \forall j \in \mathcal{C}_{IM}^+, \forall ij \in I^+$, then the optimal green proportions ($g_i^{\star t}$) in $P_{NHB\ DTA\text{-}SC}$ are determined by the remainder occupancy capacities at the downstream intersection merge cells and $\sum_{\forall i \in \Gamma^{-1}(j)} y_{ij}^{\star t} = \sum_{\forall i \in \Gamma^{-1}(j)} g_i^{\star t} Q_i^t = \delta_j^t(N_j^t - x_j^t), \forall i \in \mathcal{C}_I^+$ holds.*

Proof. We know the KKT condition related to the signal control decision variable is

$$\pi^{\star t}_{(z,p)} = Q^t_i \mu^{\star 3t}_{ij} - 1, \ \forall i \in \mathcal{C}^+_I, \ \forall ij \in \mathcal{E}_{IM}, \ \forall i \in (z,p), \ \forall t \in \mathcal{T}. \tag{4.79}$$

By combining (4.78) in Definition 4.5.5 and the complementary slackness condition (4.52) one can deduce $\pi^{\star t}_{(z,p)} = 0$. This simplifies (4.79) to

$$Q^t_i \mu^{\star 3t}_{ij} = 1 \Rightarrow \mu^{\star 3t}_{ij} = \frac{1}{Q^t_i} > 0. \tag{4.80}$$

By combining this value of $\mu^{\star 3t}_{ij}$ with the complementary slackness condition (4.51) we get

$$y^{\star t}_{ij} = g^{\star t}_i Q^t_i. \tag{4.81}$$

If we sum over all the traffic movements coming from the intersection cells to the intersection merge cell j where $\forall i \in \mathcal{C}^+_I, j \in \mathcal{C}^+_{IM}, \forall ij \in I^+$ then (4.81) can be written as

$$\sum_{\forall i \in \Gamma^{-1}(j)} y^{\star t}_{ij} = \sum_{\forall i \in \Gamma^{-1}(j)} g^{\star t}_i Q^t_i. \tag{4.82}$$

By combining (4.82) with (4.77) one can deduce

$$\sum_{\forall i \in \Gamma^{-1}(j)} y^{\star t}_{ij} = \sum_{\forall i \in \Gamma^{-1}(j)} g^{\star t}_i Q^t_i = \delta^t_j (N^t_j - x^{\star t}_j), \tag{4.83}$$

and the Proposition 4.5.6 holds. □

4.6 Numerical Solutions of the NHB DTA-SC Model

In this section, we compare solutions obtained from the NHB DTA-SC and the HB DTA-SC formulations, and present results of the different optimal control structures established earlier in Section 3.

The results are generated using a network shown in Figure 4.5 based on CTM with cell length of 60 seconds in free flow speed. The physical parameters of the network are summarized in Table 4.3. Note that the free-flow speed in the network is 15.24 meter per second (m/s) and backward propagation speed is 7.6 m/s, which results

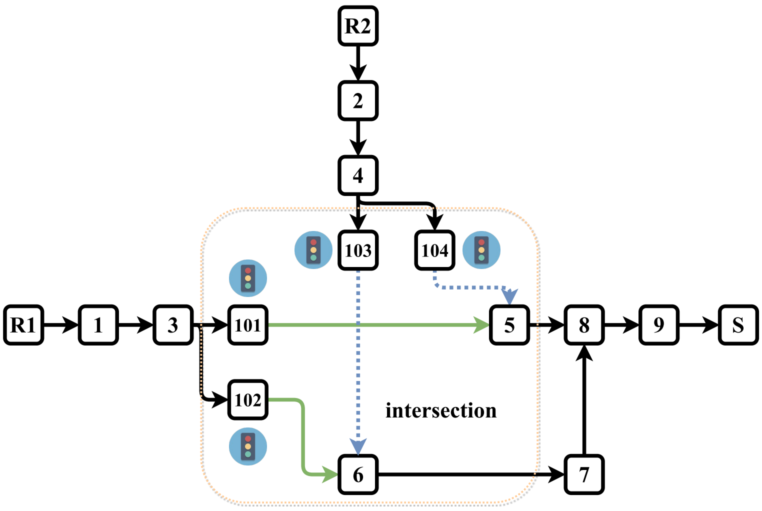


Figure 4.5. The 60s cell network under investigation. In the figure, R1 and R2 are origin or source cells and S is the sink cell. Cell 5 and 6 are the intersection merge cells. The intersection cells are marked by the traffic signal control sign. The traffic movements from cell 101 and 102 form phase-1. Whereas, movements from 103 and 104 form phase-2.

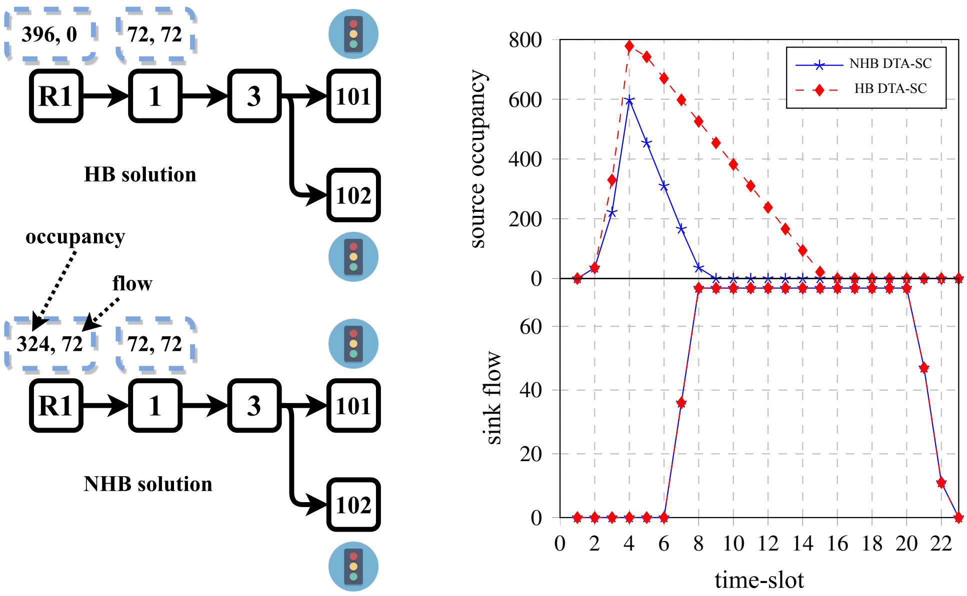
in $\delta_i = 0.5$ for all the cells. For each individual cell, the maximum flow rate Q_i and jam density N_i are presented in the Table 4.3. The generated demands at the source cells R1 and R2 are $\{90, 180, 270\}$ and $\{90, 150, 250\}$ vehicles arriving per time-slot in the first three time-slots, respectively. The optimization was solved for a time horizon of 23 time-slots or 23 minutes. Note that different demands are generated to analyze queue spillback and under-saturated scenarios. Details are provided in the corresponding result sections. The penalty levels (Ψ_i) for all the cells are also given in Table 4.3. Observe that the penalty levels are decreasing with increasing distance from the source cells. This characteristic of the penalty levels resolves the vehicle Holding-Back problem.

4.6.1 Holding-Back vs Non-Holding-Back Solution

Figure 4.6(a) shows the optimal cell occupancies and flows for both the Holding-Back SO-DTA and NHB SO-DTA solutions. Observe that the HB formulation holds 72 vehicles at $t = 4$ at the source cell R1 and it has occupancy of 396 vehicles. At this point, cell 1 has available occupancy capacity for another 72 vehicles and vehicles

Table 4.3. The physical parameters of the CTM network.

	Parameters			
Cell type	Cell index	Q_i (veh per slot)	N_i (veh per slot)	Ψ_i
Source	R1, R2	72	$+\infty$	9
Ordinary	1, 2	72	216	8
Merge	3, 4	72	216	7
Intersection	101, 104	36	108	5
Intersection	102, 103	36	108	6
Intersection merge	5	50	150	4
Intersection merge	6	50	150	5
Ordinary	7	50	150	4
Merge	8	100	300	3
Ordinary	9	100	300	2
Sink	S	$+\infty$	$+\infty$	1



(a) Cells R1 and 1 at $t = 4$.

(b) Combined occupancy of the two origins and network throughput.

Figure 4.6. (a) The solution of the HB DTA-SC and NHB DTA-SC cases at the time-slot $t = 4$ for the cells R1 and 1. In the figure, the occupancies (e.g., 324) and flows (e.g., 72) are presented in the dotted boxes as pairs. (b) Total occupancies of the source cells are presented in the upper part of the figure and the number of vehicles reaching the destination sink cell with each time-slot is presented at the bottom part of the figure. Both the formulations attain TSTT of 12185 vehicle-minute.

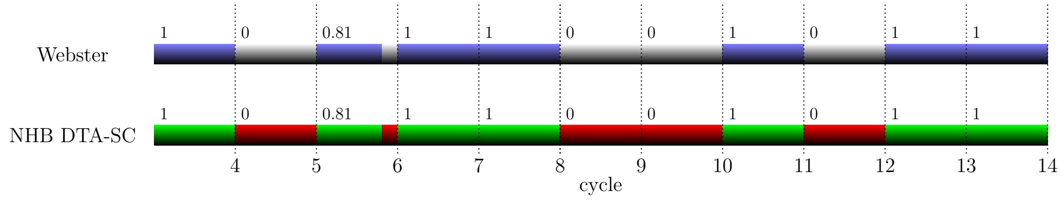


Figure 4.7. Figure showing optimal green-splits (e.g., 0.81) attained using the NHB DTA-SC in $P_{NHB\ DTA\text{-}SC}$ and NHB DTA with Webster's control for the time-slots 4 to 14 in an over-saturated traffic condition. The green-splits are presented in numbers on the bars for the mentioned traffic condition. The white, and red colors are indicative of the red-signal. Whereas, the blue, and green portions of the bars are the measures of allocated green-splits within that specific signal control cycle.

should not have been held back at the source cell R1. On the other hand, in the NHB solution those 72 vehicles are moved to cell 1 and there was no Holding-Back problem present.

In Figure 4.6(b), the plots at the top graph represents a total number of vehicles occupying the source cells R1 and R2 per time-slot. Whereas, the bottom graph shows throughputs or flows exiting the network per time-slot. It can be seen that both the HB DTA-SC and NHB DTA-SC produce the same throughput. As a result, they attain same Total System-wide Travel Time (TSTT) of 12185 vehicle-minute. Similar to Figure 4.6(a), the occupancy graph at the source indicates a large number of vehicles held back in the HB DTA-SC solution while the network operates nearly at Q_i at each cell within the network. Hardly any congestion develops within the network in the HB solution. This kind of traffic propagation is unrealistic and makes the HB formulation unusable for real-life transportation planning and applications, such as the incorporation of signal control studied in this chapter.

4.6.2 Signal Control Structure: Over-saturated Condition

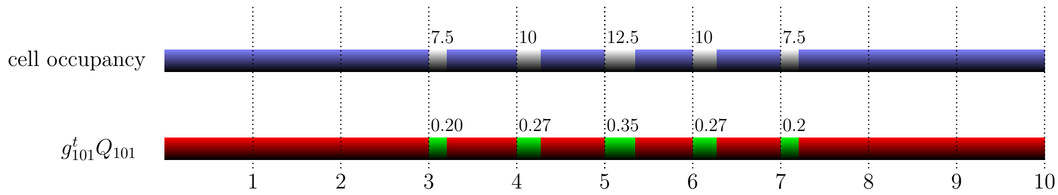
The theoretical structure of the signal control in an over-saturated condition was discussed in the Proposition 4.5.2. Herein we show a numerical example of this structure below.

We define the phase constraints such that cell 101 and cell 102 are in phase-

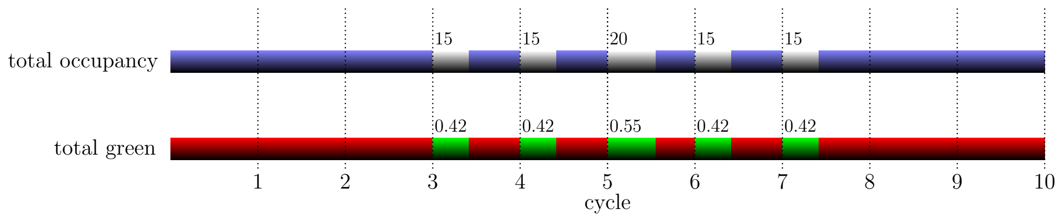
1 and get the same amount of green-splits. Whereas, cell 103 and 104 form the second phase. The cycle-length is equal to the discrete time interval (i.e., 60s). The minimum green time constraint is ignored throughout this study. The network is over-saturated because of very high input demands. We solve the same scenario for the NHB DTA-SC formulation and NHB DTA with the Webster's control policy (Webster 1958). The optimal signal splits for the cell 101 are shown in Figure 4.7 for both the solutions. Observe that the NHB DTA-SC formulation and that with Webster's control constraints formulation attain identical optimal control when the network is over-saturated during the time-slots 4 to 14, which validates the Proposition 4.5.2.

4.6.3 Signal Control Structure: Under-saturated Condition

In this section, the numerical optimal control solution in an under-saturated condition is discussed. The solution is generated using the same network in Figure 4.5. However,



(a)



(b)

Figure 4.8. Occupancies and green-splits are presented in numbers on the bars for the under-saturated condition. The white, and green colors represent occupancy and green-proportion within the time-slot. The red portions of the total green and $g^t_{101}Q_{101}$ bars are indicative of the red-signal. The blue portions of the occupancy bars are the measures of how many vehicles occupying intersection cells. (a) Cell 101 occupancy (white color) and allocated green-proportion (green color) at different time-slots are shown. (b) Summation of occupancies of the intersection cells 101 (in phase-1) and 104 (in phase-2) at different cycles (or time-slots) and total allocated green-proportions allocated to the corresponding phases are shown.

small input demands are used to create an under-saturated network condition. In particular, the demands are {15, 20, 25, 20, 15} and {15, 10, 15, 10, 15} vehicles arriving at the source cell R1 and R2, respectively. The optimal flows and green-splits are presented in Figure 4.8. Figure 4.8(a) shows that occupancies of the cell 101 and green-times are equal. Intuitively, the optimal flow has to be equal to the optimal occupancy in the under-saturated network. During the first three time-slots, the flows are zero because the generate demands at the source cells arrive at the intersection at the time-slot 4. After the 7th time slot, the flows become zero again, because, all the traffic have already left the intersection by this time. In Figure 4.8(b), it can be seen that the sum of the green-splits of the two phases at each of the time-slots does not add up to 1 as there is not enough traffic to utilize the full cycle. As the traffic demand increases the cell occupancy increases and so as the green-split which brings the optimal green-split to be proportional to the corresponding cell's optimal occupancy in an under-saturated traffic condition. This verifies the optimal control structure presented in Proposition 4.5.4.

4.6.4 Signal Control Structure: Queue Spillback Condition

Queue spillback condition arises when the remainder occupancy capacity at the intersection merge cells restricts outgoing flows from the upstream intersection cells. To create this traffic scenario, we use the same demand and network settings as in over-saturated condition and at the same time reduce the saturation flow Q_8 and occupancy capacity N_8 of cell 8 to 50 and 150, respectively. This reduced capacity of cell eight also increases TSTT to 15052 vehicle-minute.

Figure 4.9 shows the sum of optimal flows and green-proportions for the intersection cell 101 and 104 for the time-slots 8 to 17. The first seven-time slots are not shown in the figure as queue spillback only starts after this point. The remaining occupancy capacity at the intersection merge cell 5 is also shown in the figure. During the time-slots 9 to 12, the cell 5 can accommodate maximum 36 vehicles.

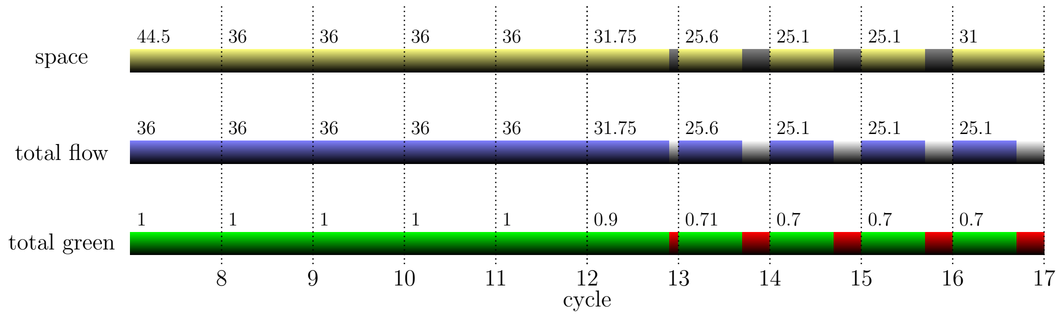


Figure 4.9. The top most bar shows total space available at the intersection merge cell 5. Uniform yellow color of the bar within a cycle means that the available occupancy capacity at cell 5 is not restricting to allocate the total green-proportion to the upstream intersection cells 101 and 104. In contrary, if the bar is divided into yellow and gray colors within a time-slot then the remainder space at cell 5 is determining the green-splits. Occupancy capacity and green-proportions are presented in numbers on the bars. The red and white portions of the other two bars indicate red-signal and flow capacity not being utilized.

Similarly, the maximum number of vehicles at cell 101 and 104 that can be served by the intersection is 36. As they are equal, the space constraint in Equation 4.30 is not restricting to allocate the total green-time to the phases. This also implies that the optimal flows out of cell 101 and 104 are also not restrained by the space constraint in cell 5. As a result, we see in the figure that the total optimal flow, accommodation capacity, and green-proportions are equal to each other during these time-slots. From the 13^{th} till 16^{th} time-slot, the available space capacity at the intersection merge cell 5 (e.g., 25.6 veh per slot) goes below the maximum green-proportion (i.e., 36 veh per slot), that can be allocated to the movements from cell 101 and 104. As a result, the space constraint at cell 5 restricts the incoming flows from the intersection cells 101 and 104. During these time-slots, we see that the total green-proportions allocated to the cell 101 and 104 are equal to the remainder space of the cell 5. This optimal solution example verifies the optimal control structure presented in Proposition 4.5.6 for the spillback traffic scenario.

4.7 Chapter Summary

In contrast to the Holding-Back traffic assignment, the Non-Holding-Back (NHB) formulation provides more realistic details about traffic propagation. This chapter discusses optimal solution structures related to flow and occupancy of an NHB SO-DTA formulation.

We have demonstrated by a thorough mathematical procedure that the presented penalty based NHB SO-DTA formulation can attain multiple optimal solutions by interchanging the occupancies within the same penalty level while keeping the optimal aggregated flows unchanged. Then we prove that the aggregated flow within a penalty level at any time-slot is always maximum attainable flow. We combine these two structures to derive that at any time-slot the NHB SO-DTA formulation minimizes delay at the each of the penalty level. Afterward, this property has been used to show that the formulation also minimizes overall network-wide delay.

Finally, we have formulated an NHB Intersection Delay Minimization (IDM) framework. The solution of this formulation is compared with that of NHB SO-DTA formulation. We have rigorously proven that the NHB SO-DTA framework also minimizes aggregated intersection delay. The numerical results show exact convergence with the presented theoretical structures. These structures provide better insights about the Non-Holding-Back traffic assignment and significant contributions in the field.

We also study the mathematical structures of the resulting optimal solution of the signal control. We propose a decomposition of NHB DTA-SC using the Lagrangian dual method where the formulated NHB DTA-SC problem is divided into three different sub-problems of occupancy minimization, flow maximization, and traffic signal control. Applying the KKT optimality conditions of the decomposed signal control sub-problem, we successfully deduce the local structures of the optimal signal control for under-saturated, over-saturated, and queue spillback traffic conditions.

The key findings are, the local structure of the optimal green splits in an over-saturated traffic situation follows the Webster's control Webster (1958); while in an under-saturated condition it is proportional to the occupancy at the intersection as in the proportional control policy of Le et al. (2015); and in a queue spillback traffic scenario the locally optimal control settings are determined by the remaining available capacity of the downstream link. Finally, we show several numerical results confirming these theoretical findings of the local optimal control structures.

Chapter 5

Bus Rapid Transit with Transit Signal Priority (BRT-TSP)

5.1 Introduction

Every year larger numbers of people are moving to urban areas. With the rising population density in cities, traffic congestion is also increasing. This has been a compelling issue for the past few decades. It is well known that one of the most promising strategies to reduce congestion on the urban roads is the BRT-TSP system (Robert 2013; Deng and Nelson 2011). Furthermore, this would substantially reduce vehicle discharged emissions and fuel consumptions (Ernst 2005; Satiennam, Fukuda, and Oshima 2006). A computationally tractable combined traffic assignment with *Signal Control (SC)* formulation would provide us a framework that can be used not only to design and analyze BRT-TSP but also to determine the optimal route choice for all the vehicles within the traffic network (Cheung and Shalaby 2017). However, most of the TSP formulations that consider *Dedicated Bus-only Lanes with Priority (DBLP)* do not consider a DTA in their formulations. On the other hand, the majority of the DTA studies that formulate BRT do not include TSP SC model in their formulations and suffer from the unwanted vehicle *Holding-Back (HB)* problem. Additionally, almost all the presented formulations are mixed-integer or non-linear models which are computationally expensive.

In summary, none of the existing methods in the literature study the combination of all three components of the BRT-TSP system, i.e., signal control, DBLP, and multi-class traffic behavior in a computationally tractable framework.

To address the above-mentioned shortcomings of the existing DTA models related to BRT and TSP, we propose a *Cell Transmission Model (CTM)* based *Bus Priority System Optimal Dynamic Traffic Assignment with Signal Control (BP SO-DTA-SC)* mathematical framework to design and analyze the combined BRT-TSP systems. This formulation includes bus routes. In addition, our model supports the give-way behavior between cars and buses at the diverging, merging, and intersection nodes as well as assures fairness to all the bus and car passengers. We also introduce a control scheme namely *Transit Priority Enabling Signal Control (TPE-SC)* model which enables TSP through the objective function and eliminates intra-phase and inter-phase conflicts at the intersections. To the best of our knowledge, this is the only SO-DTA framework that combines BRT and TSP in a *linear* formulation and can be applied to analyze city-size BRT-TSP systems.

The remainder of the chapter is organized as follows. Section 5.2 presents the novel linear-continuous BP SO-DTA-SC formulation. Our numerical results and analysis of the optimal solutions are discussed in Section 5.3. Finally, Section 5.4 summarizes the key outcomes of this chapter.

5.2 Model Formulation

In this section, we first propose the objective function that provides priority for the buses on the links and at the intersections. We also develop CTM based diverging, merging, source, sink, and intersection cells. These newly proposed cells consider Dedicated Bus-only Lanes (DBLs) and car lanes. Note that, in our model, there are two kinds of cells: (i) bus cells and (ii) car cells. Buses only run in bus cells and cars only run in car cells. This is to model BRT where a bus cell can have different saturation flow, backward propagation speed, and jam density than a car cell. These models provide a basis to formulate BRT-TSP in a multi-class traffic environment mathematically. These cell structures cover the full range of flow-density relation of the fundamental diagram and do not violate any CTM condition. As a result, they can capture real-life characteristics of traffic propagation such as congestion, queue-spillback, backward propagation, etc. Generic network topologies can be implemented using these cells. Later, we present a signal control model that enables us to implement the TSP. Finally, we present the overall linear framework, namely, BP SO-DTA-SC formulation. This formulation simultaneously guarantees NHB solutions and minimizes Total System-wide Passenger Travel Time (TSPT). The notations used throughout this chapter are outlined below.

Sets and indices

T	optimization time horizon
$\mathcal{T}$	set of all the discrete time intervals t
B	bus index
G	car index
ℓ	vehicle class, i.e., B or G
$\mathcal{C}_{\ell}$	set of all the bus (i.e., $\ell = B$) or car (i.e., $\ell = G$) cells
$\mathcal{C}_{D_\ell}$	set of all the diverging bus or car cells
$\mathcal{C}_{M_\ell}$	set of all the bus or car merging cells
$\mathcal{C}_{O_\ell}$	set of all the bus or car ordinary cells
$\mathcal{C}_{S_\ell}$	set of all the bus or car sink cells
$\mathcal{C}_{R_\ell}$	set of all the bus or car origin cells
$\mathcal{C}_{I_\ell}$	set of all the bus or car intersection cells
$\mathcal{E}_{R_\ell}$	set of $\mathcal{C}_{R_\ell}$ connectors
$\mathcal{E}_{O_\ell}$	set of $\mathcal{C}_{O_\ell}$ connectors
$\mathcal{E}_{M_\ell}$	set of $\mathcal{C}_{M_\ell}$ connectors
$\mathcal{E}_{D_\ell}$	set of $\mathcal{C}_{D_\ell}$ connectors
$\mathcal{E}_{S_\ell}$	set of all the bus or car sink cell connectors
$\mathcal{E}_{S}$	$\mathcal{E}_{S_B} \cup \mathcal{E}_{S_G}$
$\Gamma^{-1}(i_\ell)$	set of all the predecessor cells of the cell i_ℓ
$\Gamma(i_\ell)$	set of all the successor cells of the cell i_ℓ
$\mathcal{I}$	set of all the intersections where I is an intersection $\in \mathcal{I}$
P	set of all the phases
z	intersection index
n	is an artificial index to label incoming flows

Parameters

τ	length of each time-slot
$v_{i_\ell}^t$	free flow speed at cell $i_\ell, \ell \in \{B, G\}$ at time-slot t
$w_{i_\ell}^t$	backward speed at cell i_ℓ at time-slot t
$\delta_{i_\ell}^t$	the ratio of $v_{i_\ell}^t / w_{i_\ell}^t$
$d_{i_\ell}^t$	deterministic demand at the source cell i_ℓ at time-slot t
$N_{i_\ell}^t$	occupancy capacity of the cell i_ℓ at the time-slot t
$Q_{i_\ell}^t$	maximum flow into/out of cell i_ℓ at time-slot t
Ψ_{i_B}	penalty on bus occupancy at cell i_B
Ψ_{i_G}	penalty on car occupancy at cell i_G
α_B	average number of passengers per bus
α_G	average number of passengers per car
$f_{i_B j_B}^{R_B}$	proportion of the demand generated at a bus source cell R_B that must go through cell i_B

In the above list, time-slot duration τ is a CTM parameter. Based on τ, the continuous assignment period of interest is discretized into number of small time-slots (Ziliaskopoulos 2000). In CTM, cell length $L_\ell = \tau \cdot v_{i_\ell}^t$ (Daganzo 1994). As a result, the time-slot duration and free flow speed combinedly determine the cell length. The CTM parameter $N_{i_\ell}^t$ is the maximum number of vehicles that can be accommodated by the cell i_ℓ at time-slot t. If a cell has jam density of $K_{jam_{i_\ell}}^t$, then $N_{i_\ell}^t = K_{jam_{i_\ell}}^t \cdot L_\ell$. Therefore, the $v_{i_\ell}^t$ also influences the occupancy capacity. For a triangular fundamental diagram of the CTM and known values of free flow speed ($v_{i_\ell}^t$), backward propagation speed ($w_{i_\ell}^t$), and jam density ($K_{jam_{i_\ell}}^t$), the maximum inflow/outflow capacity of a cell i_ℓ at time-slot t is $Q_{i_\ell}^t = \frac{K_{jam_{i_\ell}}^t \cdot v_{i_\ell}^t \cdot w_{i_\ell}^t \cdot \tau}{v_{i_\ell}^t + w_{i_\ell}^t}$. The $w_{i_\ell}^t$ is always less than or equal to the $v_{i_\ell}^t$. As a result, their ratio $\delta_{i_\ell}^t \leqslant 1$. A commonly used value of $\delta_{i_\ell}^t$ is 0.5. The demand parameter $d_{i_\ell}^t$ is the number of vehicles arriving at a source cell i_ℓ at time-slot t. We consider predetermined demands that are input to the proposed DTA model. Rest of the parameters in the above list are discussed in the following sections.

Decision variables

$x^t_{i_\ell}$	occupancy in cell $i_\ell, \ell \in \{B, G\}$ at time-slot t
$y^t_{i_\ell j_\ell}$	flow from cell i_ℓ to j_ℓ, $\ell \in \{B, G\}$ at time-slot t
$\Omega^t_{(z,p)_\ell}$	phase green-split allocated to cell $(z, p)_\ell, z \in \mathcal{I}, p \in P$ at time-slot t
$\omega^t_{(z,p)_\ell}$	dis-aggregated green-split allocated to cell $(z, p)_\ell$ at time-slot t
$b^t_{i_\ell, j_\ell n}$	optimal give-way proportion allocated to flow $y_{i_\ell, j_\ell n}$ at time-slot t at a diverging or a merging cell

5.2.1 Objective Function of the BP SO-DTA Framework

In this section, we propose the objective function of the BP SO-DTA formulation. This objective function has the following three key properties:

1. It eliminates the HB problem.
2. It considers multiple traffic-classes.
3. It gives priority to the buses over cars.

The proposed NHB objective function of the BP SO-DTA model is provided below.

$$\mathcal{O} = \min \sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C}_B \setminus \mathcal{C}_{S_B}} \alpha_B \cdot \Psi_{i_B} \cdot x^t_{i_B} + \sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C}_G \setminus \mathcal{C}_{S_G}} \alpha_G \cdot \Psi_{i_G} \cdot x^t_{i_G}. \tag{5.1}$$

In Equation (5.1), B is the subscript for bus, and G is the subscript for car. Let T denote the optimization time horizon, and t be the time-slot index. Parameter α_B is the average number of people occupying a bus and α_G is that of car. The decision variables $x^t_{i_B}$ and $x^t_{i_G}$ are the number of bus and car vehicles occupying cell i at time-slot t, respectively. As a result, the objective function minimizes overall occupancy of the network where the predefined penalty terms Ψ_{i_B}s are associated with bus cells, and Ψ_{i_G}s are associated with car cells.

In the algorithm presented in Section 5.2.1.1 we note that $\Psi_{i_B} = \Psi_{i_G}$ if they are at the same relative distance from the destination cell. We prove later in Proposition B.2.1 that the vehicle HB problem can be eliminated by multiplying a cell-dependent penalty with the occupancy of that cell. The highest penalty is associated

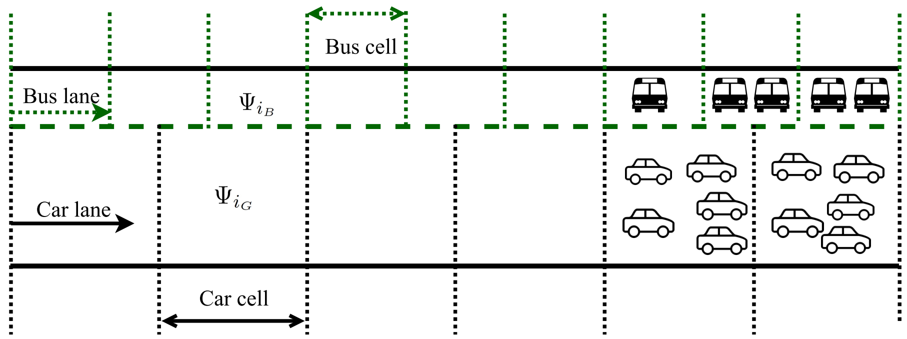


Figure 5.1. An example of road segment consisting of DBL and car lanes which are then divided into several car and bus cells. Different penalty terms of the cells, i.e., Ψ_{i_B}, Ψ_{i_G} are associated with their distances from the destination.

with the source cells, and the penalty decreases as one moves closer towards the destination. If cell i is upstream to cell j, then $\Psi_{i_B} > \Psi_{j_B}$ for the bus occupancy and $\Psi_{i_G} > \Psi_{j_G}$ for the car occupancy. Furthermore, $0 < \Psi_{i_B} \leq 1$ and $0 < \Psi_{i_G} \leq 1$.

The parameters α_B and α_G work as scaling factors for the penalty terms so that in Equation (5.1) maximum possible value of $\alpha_B \cdot \Psi_{i_B} \leq \alpha_B$ and $\alpha_G \cdot \Psi_{i_G} \leq \alpha_G$. By setting $\alpha_B > \alpha_G$ we implement the bus priority, i.e., the cost of delaying a transit vehicle is more expensive than that of a car. Therefore, the higher incentive in the objective function to serve buses at any links necessarily models the bus-priority property of the BRT. An example scenario is provided in Figure 5.1. In the figure, the road is divided into two lanes consisting of the BRT lane and the car lane. The lanes are further divided into the corresponding classes of cells, i.e., separately bus and car cells. In the figure, the terms Ψ_{i_B} and Ψ_{i_G} represent the penalty on the bus and car occupancy, respectively.

5.2.1.1 Algorithm for Determining Ψ_{i_B} and Ψ_{i_G} Penalty Terms

The algorithm provided below is to determine the penalty terms for general networks including the symmetric and cyclic ones in the multi-class traffic scenario. This algorithm is loop free and can handle different cell-lengths of bus and car cells. An example of the mentioned use-case of this algorithm is provided in A.1.

Remark 4. *Important notes related to the algorithm are enlisted below.*

- *This algorithm is designed for single destination (per vehicle class) CTM networks. In the CTM, a connector that connects a diverging and a merging cell is not allowed (Daganzo 1995; Ziliaskopoulos 2000).*
- *All the bus cells have length of L_B. Similarly, all the car cells have homogeneous cell length L_G. But L_B and L_G may not be equal.*
- *Here i_B and j_B are bus cell indexes, and $i_B \in \Gamma^{-1}(j_B)$. Similarly, i_G and j_G are car cell indexes where $i_G \in \Gamma^{-1}(j_G)$.* $\odot$

Definition 5.2.1. (Definition of path for determining Ψ_{i_B} and Ψ_{i_G} values) A path $p_{a(S_\ell \to R_\ell)}$, $\ell \in \{B, G\}$ is an ordered list of cells starting from the sink cell $\mathcal{C}_{S_\ell}$ and goes towards a source cell $\mathcal{C}_{R_\ell}$ by facing each of the cell connector (directional) from the opposite direction where $a(S_\ell \to R_\ell)$ is the path-index. A bus path consists of bus cells only (i.e., when $\ell = B$) and a car path contains car cells only (i.e., when $\ell = G$). If a path

- $a(S_\ell \to R_\ell)$ consists of no cell more than once, then the path ends at a source cell R_ℓ;
- $a(S_\ell \to R_\ell)$ repeats a cell j_ℓ for the second time, then the path ends at cell i_ℓ.

$\triangle$

Start

1. Initialize:

   (a) Determine L_B and L_G.

   (b) Set $\Psi_{S_B} = 0$; $\Psi_{S_G} = 0$.

(c) Determine set of all the bus paths $\mathcal{P}_B$ and car paths $\mathcal{P}_G$ in which each path follows Definition 5.2.1.

2. Determine path-based penalty terms for the bus and car cells using the following relations. Start from predecessor-cell of the sink cell and move towards the source cell. Stop when a path ends (i.e., the last cell in the path). Repeat for other paths.

   (a) For $\forall p_{a(S_B \to R_B)} \in \mathcal{P}_B$: $\Psi_{i_B p_{a(S_B \to R_B)}} = \Psi_{j_B p_{a(S_B \to R_B)}} + L_B$, when, $i_B \notin \mathcal{C}_{S_B} \cup \mathcal{C}_{R_B}$.

   (b) For $\forall p_{a(S_G \to R_G)} \in \mathcal{P}_G$: $\Psi_{i_G p_{a(S_G \to R_G)}} = \Psi_{j_G p_{a(S_G \to R_G)}} + L_G$, when, $i_G \notin \mathcal{C}_{S_G} \cup \mathcal{C}_{R_G}$.

   (c) For $\forall p_{a(S_B \to R_B)} \in \mathcal{P}_B$: $\Psi_{i_B p_{a(S_B \to R_B)}} = \Psi_{j_B p_{a(S_B \to R_B)}} + \min(L_B, L_G)$, when, $i_B \in \mathcal{C}_{R_B}$.

   (d) For $\forall p_{a(S_G \to R_G)} \in \mathcal{P}_G$: $\Psi_{i_G p_{a(S_G \to R_G)}} = \Psi_{j_G p_{a(S_G \to R_G)}} + \min(L_B, L_G)$, when, $i_G \in \mathcal{C}_{R_G}$.

3. Convert all the path specific penalty terms to cell specific penalty terms using following relations, i.e., if a cell has two or more path specific penalty terms, then select their maximum value; if a cell has just one path specific penalty term then take it as the cell specific penalty.

   (a) $\Psi_{i_B} = \max(\Psi_{i_B p_{a(S_B \to R_B)}})$, $i_B \notin \mathcal{C}_{S_B}$, $\forall p_{a(S_B \to R_B)} \in \mathcal{P}_B$.

   (b) $\Psi_{i_G} = \max(\Psi_{i_G p_{a(S_G \to R_G)}})$, $i_G \notin \mathcal{C}_{S_G}$, $\forall p_{a(S_G \to R_G)} \in \mathcal{P}_G$.

4. Determine Ψ_{max}, which is the maximum of all the penalty terms among all the bus and car cells in the network, i.e., $\max(\Psi_{i_\ell}, \forall \ell \notin \mathcal{C}_{S_\ell})$. Then normalize the values of Ψ_{i_B} and Ψ_{i_G} determined in the step-3 as follows:

   (a) $\Psi_{i_B} = \frac{\Psi_{i_B}}{\Psi_{max}}$, $\forall i_B \notin \mathcal{C}_{S_B}$.

(b) $\Psi_{i_G} = \frac{\Psi_{i_G}}{\Psi_{max}}, \ \forall i_G \notin \mathcal{C}_{S_G}$.

End

In the following sections, different proposed cell structures enable us to assign the bus and car vehicles within the same optimization framework. These models can be applied to obtain solutions for any single-destination networks.

5.2.2 Ordinary Cell Model

An ordinary cell has only one successor cell and one predecessor cell. In the BP SO-DTA formulation, the bus and car lanes are divided into several bus and car cells (see Figure 5.1), respectively. The bus and the car cells can have different free-flow speeds, backward-speeds, saturation-flows, and jam densities. The separation of the bus and car lanes enables us to establish cells separately for bus and car traffic. The advantage of this modeling approach is two-fold. Firstly, the parameters and length of a bus cell are independent of car cells. Secondly, the bus vehicle class does not impose the moving bottleneck problem (Liu et al. 2015a). Therefore, this enables us to avoid computationally expensive mixed-integer or nonlinear programming. The linear equality and inequality constraints that constitute a bus ordinary cell and a car ordinary cell are presented below.

$$x^t_{i_\ell} = x^{t-1}_{i_\ell} + y^{t-1}_{k_\ell i_\ell} - y^{t-1}_{i_\ell j_\ell}, \tag{5.2}$$

$$y^t_{i_\ell j_\ell} \leq x^t_{i_\ell}, \tag{5.3}$$

$$y^t_{i_\ell j_\ell} \leq \min(Q^t_{i_\ell}, Q^t_{j_\ell}), \tag{5.4}$$

$$y^t_{i_\ell j_\ell} \leq \delta^t_{j_\ell}(N^t_{j_\ell} - x^t_{j_\ell}), \tag{5.5}$$

where $\forall \ell \in \{B, G\}, \forall k_\ell \in \Gamma^{-1}(i_\ell), \ \forall i_\ell \in \mathcal{C}_{O_\ell}, \ \forall j_\ell \in \Gamma(i_\ell),$

$\forall i_\ell j_\ell \in \mathcal{E}_{O_\ell}, \ \ \forall t \in \mathcal{T}.$

In the above, ℓ is the vehicle-class index where B and G stand for the bus and car traffic classes, respectively. Equation (5.2) is the flow-occupancy balance relation.

Constraint (5.3) ensures that the flow does not exceed occupancy at cell i_ℓ at any time t. Constraints (5.4) and (5.5) restrict the outgoing flow of cell i_ℓ by the saturation flow (equal to $\min(Q^t_{i_\ell}, Q^t_{j_\ell})$) and the remaining occupancy capacity of the successor cell, respectively. If one wants to model the bus and car lanes within the same cell, the above constraints would be similar. However, the saturation flow capacity (the remaining occupancy capacity) of a car lane would be determined by subtracting the bus saturation flow (the bus occupancy capacity) from the total flow capacity (the total occupancy capacity). In that case, the free-flow speed has to be equal for both the bus and the car. To overcome this limitation, we have used separate classes (e.g., different speed, capacity) of cells for the bus and the car lanes.

5.2.3 Diverging Cell Model

Our proposed diverging cells for the diverging BRT and car lane have a different structure than that of the *Holding-Back System Optimal Dynamic Traffic Assignment (HB SO-DTA)* formulation proposed by Ziliaskopoulos (2000). As shown in Figure 5.2, there is a conflict caused by the traffic movements from the bus and car diverging cells. The mathematical model that forms the bus and the car diverging cells is outlined below.

$$x^t_{i_\ell} = x^{t-1}_{i_\ell} + y^{t-1}_{k_\ell i_\ell} - \sum_{j_\ell \in \Gamma(i\ell)} y^{t-1}_{i_\ell j_\ell n}, \tag{5.6}$$

$$\sum_{j_\ell \in \Gamma(i_\ell)} y^t_{i_\ell j_\ell n} \leq x^t_{i_\ell}, \tag{5.7}$$

$$\sum_{j_\ell \in \Gamma(i_\ell)} y^t_{i_\ell j_\ell n} \leq Q^t_{i_\ell}, \tag{5.8}$$

$$y^t_{i_\ell j_\ell n} \leq \delta^t_{j_\ell}(N^t_{j_\ell} - x^t_{j_\ell}), \tag{5.9}$$

where $\forall \ell \in \{B, G\},\ \forall k_\ell \in \Gamma^{-1}(i_\ell),\ \forall j_\ell \in \Gamma(i_\ell),\ \forall i_\ell j_\ell \in \mathcal{E}_{D_\ell},$

$n \in \mathcal{D},\ \forall i_\ell \in \mathcal{C}_{D_\ell},\ \forall t \in \mathcal{T}.$

In the above, Equation (5.6) is the occupancy-flow conservation equation at the diverging cell. Constraints (5.7)-(5.8) control the outgoing flow of a diverging cell by

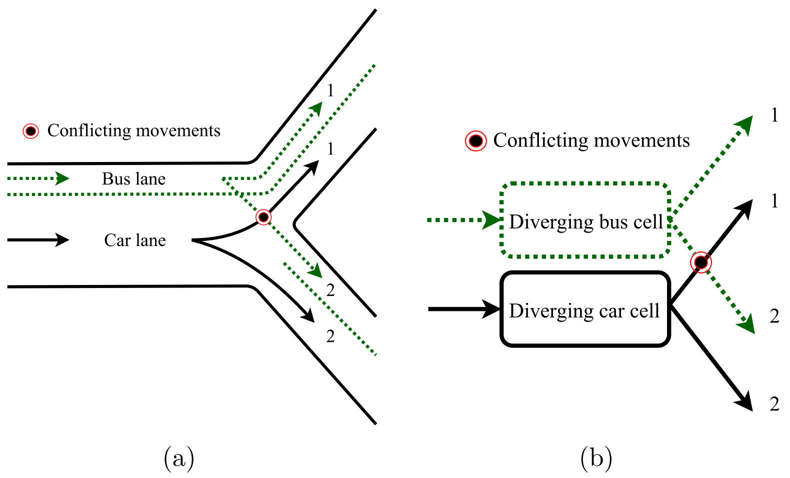


Figure 5.2. (a) An example of the diverging DBL in one incoming and two outgoing links scenario. (b) The cell representation of the diverging link and the conflicting traffic movements between the diverging bus and car cells.

the available occupancy and flow capacity of the cell, respectively. The last constraint in Equation (5.9) ensures that the outgoing flows at the diverging cell do not exceed the remainder capacities at its successor cells.

To eliminate the conflicting movements between the DBL and the car lane, we use a conflict avoidance variable $b^t_{i_B j_B n}$ for the diverging bus cell and $b^t_{i_G j_G n}$ for the diverging car cell, where n is a local-index of the outgoing connector from cell i to cell j. In the Figure 5.2, we see that there are two outgoing connectors from each of the bus and car diverging cells. These outgoing connectors have local-index of 1 and 2. In the figure, the bus movement along the bus connector 2, and the car movement along the car connector 1 conflict with each other. For a more generic formulation, we introduce new sets in which each element represents conflicting bus and car arcs. Since there is no conflict in between the very first outgoing connector of a bus diverging cell and the last outgoing connector of a neighbor car diverging cell, we define a set of conflicting bus arcs as $\mathcal{D}_B = \{2, 3, ..., e\}$ and that of car as $\mathcal{D}_G = \{1, 2, 3, ..., e-1\}$ where e is the last outgoing arc (e.g., bus connector 2 and car connector 2 in Figure 5.2). The following equations formulate the outgoing flows from the diverging cells i_B and i_G to resolve the conflicts between the bus and the car vehicle movements.

$$y^t_{i_B j_B 1} \leq Q^t_{j_B}, \tag{5.10}$$

$$y^t_{i_G j_G e} \leq Q^t_{j_G}, \tag{5.11}$$

$$y^t_{i_B j_B n} \leq b^t_{i_B j_B n} Q^t_{j_B}, n \in \mathcal{D}_B, \tag{5.12}$$

$$y^t_{i_G j_G n} \leq b^t_{i_G j_G n} Q^t_{j_G}, n \in \mathcal{D}_G, \tag{5.13}$$

$$b^t_{i_B j_B n} + b^t_{i_G j_G (n-1)} \leq 1, n \in \mathcal{D}_B, \tag{5.14}$$

$$b^t_{i_B j_B 2} = ... = b^t_{i_B j_B e}, \ \forall i_B \in \mathcal{C}_{D_B}, \ j_B \in \Gamma(i_B), \ \forall t \in \mathcal{T}, \tag{5.15}$$

$$b^t_{i_G j_G 1} = ... = b^t_{i_G j_G (e-1)}, \ \forall i_G \in \mathcal{C}_{D_G}, \ j_G \in \Gamma(i_G), \ \forall t \in \mathcal{T}. \tag{5.16}$$

As discussed before, the first bus connector and the last car connector are not in conflict with each other. As a result, the constraints (5.10) and (5.11) do not include any give-way variables and their corresponding flows are limited by their own flow capacities. In the constraints (5.12)-(5.13), the give-way variables $b^t_{i_B j_B n}$ and $b^t_{i_G j_G (n-1)}$ ensure that the conflicting movements between the bus and the car classes are resolved. Furthermore, these give-way splits are applied on the saturation flow to separate the allocated time for the each of the vehicle class. The constraint (5.14) ensures that the total allocated give-way time proportions of the buses and cars does not exceed 1 (i.e., the time-slot duration). All the conflicting outgoing traffic movements belonging to the same vehicle class should get the same allocated time, this is implemented by the constraints (5.15)-(5.16). The incoming flows to the bus and car diverging cells are determined by the respective ordinary cell constraints presented in the previous section. Due to the NHB objective function and higher penalty on holding one unit of bus vehicle, the bus vehicle will get priority at the diverging links.

5.2.4 Merging Cell Model

In this section, we propose bus and car merging cell models of the BP SO-DTA formulation. These cells also have conflicting traffic movements in between the DBL

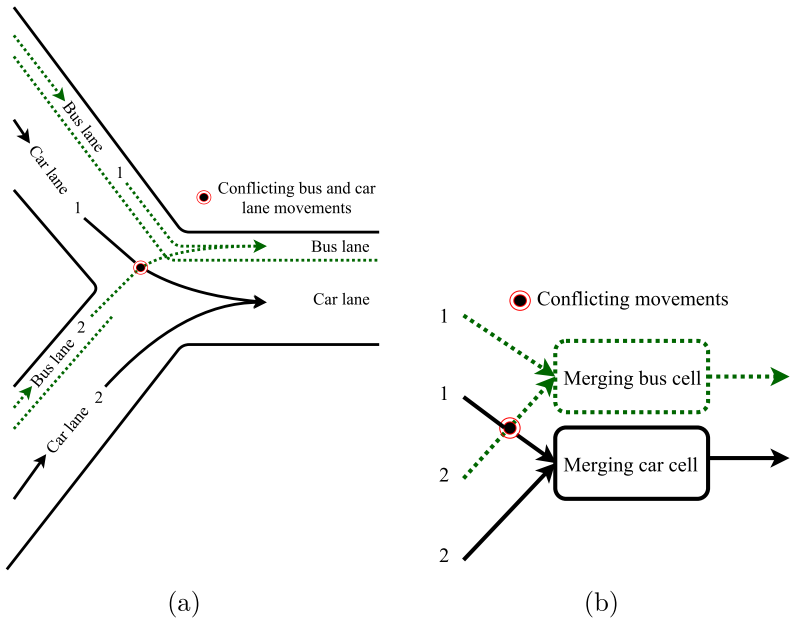


Figure 5.3. (a) The DBL in a two merging links and one outgoing link scenario. (b) The corresponding bus and car merging cells and conflicting traffic movements.

and the car lane. Thus, these cell models are different than the merging cell model presented in the HB SO-DTA formulation (Ziliaskopoulos 2000). Figure 5.3 shows the conflicting traffic movements between the buses and the cars for the two merging links. The constraints that formulate the car and bus merging cell models are described below.

$$x^t_{i_\ell} = x^{t-1}_{i_\ell} + \sum_{k_\ell \in \Gamma^{-1}(i_\ell)} y^{t-1}_{k_\ell i_\ell n} - y^{t-1}_{i_\ell j_\ell n}, \tag{5.17}$$

$$y^t_{k_\ell i_\ell n} \leq x^t_{k_\ell}, \tag{5.18}$$

$$y^t_{i_\ell j_\ell n} \leq \delta^t_{j_\ell}(N^t_{j_\ell} - x^t_{j_\ell}), \tag{5.19}$$

$$\sum_{k_\ell \in \Gamma^{-1}(i_\ell)} y^t_{k_\ell i_\ell n} \leq Q^t_{i_\ell}, \tag{5.20}$$

where $\forall \ell \in \{B, G\},\ \forall k_\ell \in \Gamma^{-1}(i_\ell),\ \forall j_\ell \in \Gamma(i_\ell),\ \ \forall k_\ell i_\ell \in \mathcal{E}_{M_\ell},$

$n \in \mathcal{M},\ \ \forall i_\ell \in \mathcal{C}_{M_\ell}, \forall t \in \mathcal{T}.$

In the above, the equality constraint (5.17) represents the occupancy-flow conservation rule in a bus or car merging cell. The inequality constraint (5.18) implies that the flow cannot exceed the occupancy of the respective cell. Constraint (5.19) ensures that the flow does not surpass the available occupancy capacity of the successor cell of the merging cell. Constraint (5.20) restricts the incoming flows to the car (bus) merging cell so that the sum of them does not exceed the saturation flow capacity of the corresponding car (bus) merging cell.

To resolve the conflicts between the incoming flows to the car and bus merging cells, we use similar give-way variables as in the previous section. We also use the conflicting connector index set for the merging cells. The conflicting movements in between the bus and car merging cells are $\mathcal{M}_B = \{2, 3, ..., e\}$ and $\mathcal{M}_G = \{1, 2, 3, ..., e-1\}$ where e is the last incoming connector. The conflict avoidance method controls the incoming flows from the predecessor cells of the merging bus and car cells. It is presented below.

$$y^t_{k_B i_B n} \leq b^t_{k_B i_B n} Q^t_{k_B}, n \in \mathcal{M}_B, \tag{5.21}$$

$$y^t_{k_G i_G n} \leq b^t_{k_G i_G n} Q^t_{k_G}, n \in \mathcal{M}_G, \tag{5.22}$$

$$b^t_{k_B i_B n} + b^t_{k_G i_G (n-1)} \leq 1, n \in \mathcal{M}_B, \tag{5.23}$$

$$b^t_{k_B i_B 2} = ... = b^t_{k_B i_B e},\ \forall i_B \in \mathcal{C}_{M_B},\ k_B \in \Gamma^{-1}(i_B),\ \forall t \in \mathcal{T}, \tag{5.24}$$

$$b^t_{k_G i_G 1} = ... = b^t_{k_G i_G (e-1)},\ \forall i_G \in \mathcal{C}_{M_G},\ k_G \in \Gamma^{-1}(i_G),\ \forall t \in \mathcal{T}. \tag{5.25}$$

In the above equations, $b^t_{k_B i_B n}$ and $b^t_{k_G i_G n}$ are the give-way decision variables to avoid conflict of incoming traffic movements from the bus and the car merging cells. The constraints (5.21) and (5.22) control incoming traffic movements to the merging cells via the optimal values of the give-way variables. These give-way splits ensure that the bus vehicle class is served first as this would provide a higher benefit to the objective function presented in Equation (5.1). These splits are optimally allocated which ensures that unnecessary delays are not suffered by any of the vehicles. Both these constraints along with (5.23) ensure that there is no conflict in between the bus and the car incoming traffic movements. Constraints (5.24) and (5.25) imply that the allocated portions of the give-way split for the same traffic class are equal.

5.2.5 Source Cell Model

The source cell model for the BP SO-DTA problem has a similar formulation to the HB SO-DTA source cell model except for the fact that the constraints are separately considered for both the bus and the car traffic classes (i.e., separate bus and car source cells). The CTM based constraints that formulate the BP SO-DTA bus and car source cell models are presented below.

$$x_{i_\ell}^{t} = x_{i_\ell}^{t-1} + d_{i_\ell}^{t-1} - y_{i_\ell j_\ell}^{t-1}, \tag{5.26}$$

$$y_{i_\ell j_\ell}^{t} \leq x_{i_\ell}^{t}, \tag{5.27}$$

$$y_{i_\ell j_\ell}^{t} \leq Q_{j_\ell}^{t}, \tag{5.28}$$

$$y_{i_\ell j_\ell}^{t} \leq \delta_{j_\ell}^{t}(N_{j_\ell}^{t} - x_{j_\ell}^{t}), \tag{5.29}$$

$$\text{where } \forall i_\ell \in \mathcal{C}_{R_\ell},\ j_\ell \in \Gamma(i_\ell),\ \forall \ell \in \{B, G\},\ \forall t \in \mathcal{T}.$$

Constraint (5.26) represents the flow occupancy balance relation in the both the bus and the car source cells. The demand parameters $d_{i_B}^{t-1}$ and $d_{i_G}^{t-1}$ enable us to set how many buses and cars enter into the network per time-slot. In this chapter, we consider time varying but deterministic demands. The constraint (5.27) restricts the outgoing flow of a source cell so that the flow cannot exceed the occupancy of it. A source cell has infinite occupancy capacity. However, it has finite outgoing flow capacity which is ensured by the constraint (5.28). This is actually imposed by the saturation flow capacity of the successor cell of the source cell. Finally, the constraint (5.29) ensures that the outgoing flow of a source cell does not exceed the remaining accommodation space at the successor cell.

5.2.6 Sink Cell Model

Similar to the HB SO-DTA sink cell, both the bus and car sink cells have infinite flow and occupancy capacity. However, the incoming flows to the sink cells are determined by the traffic state at the predecessor cells. If there are multiple predecessor cells, the conflicts between the multiple traffic movements to the sink cells can be eliminated using the same approach as the merging cells presented in the Section 5.2.4. The bus and car sink cells for the BP SO-DTA model consist of the following constraints.

$$y^t_{k_\ell i_\ell} \leq Q^t_{k_\ell}, \tag{5.30}$$

$$y^t_{k_\ell i_\ell} \leq x^t_{k_\ell}, \text{where } \forall \ell \in \{B, G\},\ k_\ell \in \Gamma^{-1}(i_\ell),\ \forall i_\ell \in \mathcal{C}_{S_\ell},\ \forall t \in \mathcal{T}. \tag{5.31}$$

In the above, the first constraint (5.30) implies that the incoming flow to a bus (car) sink cell is restricted by the bus (car) lane saturation flow of the predecessor cell. The second constraint (5.31) restricts the flow to the sink cell so that it does not exceed the available number of vehicles at the predecessor cell.

5.2.7 A Bus Route Model for the BP SO-DTA Formulation

In this section, we propose a bus route model for the bus vehicle class. The equation that formulates the pre-defined bus-route property of the buses is presented below.

$$\sum_{\forall t \in \mathcal{T}} y^t_{i_B j_B} \geq \sum_{\forall t \in \mathcal{T}} \sum_{\forall R_B \in \mathcal{C}_{R_B}} f^{R_B}_{i_B j_B} d^t_{R_B},\ \forall i_B \in \mathcal{C}_{D_B} \cup \mathcal{C}_{M_B} \cup \mathcal{C}_{O_B},\ j_B \in \Gamma(i_B). \tag{5.32}$$

The constraint (5.32) restricts the optimal out-going flow of a bus cell i_B by the proportion $f^{R_B}_{i_B j_B}$ of the total bus demand $\sum_{\forall t \in \mathcal{T}} d^t_{R_B}$ originated at a bus source cell R_B. $f^{R_B}_{i_B j_B}$ is a deterministic bus Route Parameter (RP) of the model where R_B is the index of a bus source cell. For several origins in a single destination network, if one wants that certain proportions of the demands generated at the each of the origin cells must go through a cell (in a route), then several RPs should be set corresponding to the each of the origins by using superscript R_B of the RP. By applying this constraint to several bus cells, a bus route can be defined so that a minimum number of buses go through it. The above model is based on the assumption that the buses that go through a route were actually assigned to that bus-route. This assumption is not required if one uses labeled demands (e.g., bus names) or path-based CTM that are out of the scope of this research. In the next section, the novel linear and continuous TPE-SC model is outlined.

5.2.8 The TPE-SC Model

This section presents the traffic signal control model when DBLs are introduced at intersections. The CTM constraints for the bus and car intersection cells are presented below.

$$x_{i_\ell}^t = x_{i_\ell}^{t-1} + y_{k_\ell i_\ell}^{t-1} - y_{i_\ell j_\ell}^{t-1}, \tag{5.33}$$

$$y_{i_\ell j_\ell}^t \le x_{i_\ell}^t, \tag{5.34}$$

$$y_{i_\ell j_\ell}^t \le Q_{j_\ell}^t, \tag{5.35}$$

$$y_{i_\ell j_\ell}^t \le \delta_{j_\ell}^t (N_{j_\ell}^t - x_{j_\ell}^t), \tag{5.36}$$

$$\text{where } \forall k_\ell \in \Gamma^{-1}(i_\ell),\ j_\ell \in \Gamma(i_\ell),\ \forall i_\ell \in \mathcal{C}_I,\ \forall \ell \in \{B, G\},\ \forall t \in \mathcal{T}. \tag{5.37}$$

The above relations specify the occupancy-flow balance relation and flow propagation rules with the physical parameters of a particular link. In the following, we describe the signal control model at the intersection to avoid both the inter-phase and intra-phase conflicts between the bus and car flows.

Inter-phase conflicts occur when several traffic movements in different phases attempt to pass through the intersection. To allocate green time for a phase, a signal control decision variable $\Omega_{(z,p)}^t$ is introduced. The index z specifies intersection number. On the other hand, phase index p defines which phase the variable represents. The combination of these two indices such as (z, p) indicates an intersection cell index and the respective signal control variable $\Omega_{(z,p)}^t$ controls the outgoing flow of that cell. For example, if the intersection and phase indices are 1 and 1 respectively, then the index of the intersection cell is 101 and the corresponding signal control variable is Ω_{101}^t where t is the index of the time-slot.

We also amend the signal phasing convention prescribed by the National Electrical

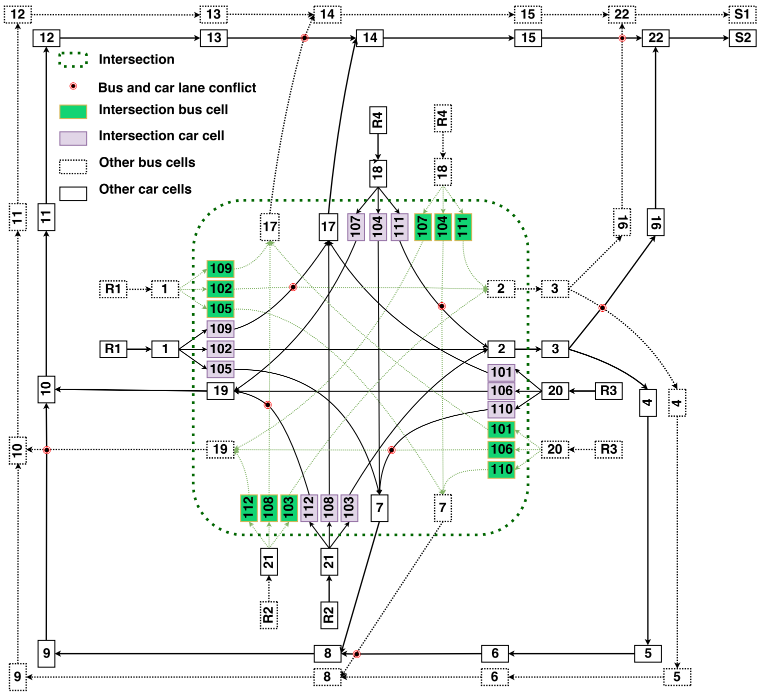


Figure 5.4. An intersection for the conflicting bus and car traffic movements is shown in the above network. Physical parameters of the network and relevant optimal solutions attained using this network are presented in Section 5.3.1.

Manufacturers Association (NEMA) for the Australian signal phasing rules. However, any other convention can be adopted using the approach presented in this chapter. The intersection model is presented in Figure 5.4. The amended signal phasing rules mandate that the phase pairs of (1,5) (i.e., cell 101 and 105 in the figure), (2,6) (i.e., cell 102 and 106), (3,7) (i.e., cell 103 and 107), and (4,8) (i.e., cell 104 and 108) are allowed together. The phase pairs (1,5) and (3,7) represent turning-right movements. Whereas, the (2,6) and (4,8) pairs form through movements. Furthermore, to grant a left turn with a through movement, the phase pairs (2,9) (i.e., cell 102 and 109), (4,11) (i.e., cell 104 and 111), (6,10) (i.e., cell 106 and 110), or (8,12) (i.e., cell 108 and 112) are permitted. The control constraints that are applied to the intersection cells are presented below. Note that in this signal control model, cycle-length is equal to the time-slot duration.

$$y^{t}_{(z,p)_\ell j_\ell} \leqslant \Omega^{t}_{(z,p)_\ell} Q_{(z,p)_\ell} - g^{o}_{(z,p)_\ell} Q_{(z,p)_\ell}, \; z \in \mathcal{I}, \; p \in \{1,2,3,4\}, \; (z,p) \in \mathcal{C}_I. \tag{5.38}$$

In Equation (5.38), $\ell \in \{B, G\},\ j_\ell \in \Gamma(i_\ell),\ t \in \mathcal{T}$. This constraint is applied to all the phases, i.e., {1,2,3,4} for both the bus and car traffic class. The deterministic quantity $g^o_{(z,p)_\ell} Q_{(z,p)_\ell}$ is the loss due to phase-switching. The typical value of $g^o_{(z,p)_\ell}$ is $\frac{1}{Q_{(z,p)_\ell}}$. For the turning right movements (i.e., phase 1 and 3) there are no intra-phase conflicting movements between the bus and the car intersection cells. Through and left movements are allowed together in the phase-2 and phase-4. In the each of these phases, four traffic movements (e.g., 2, 9, 6, 10) are allowed simultaneously. This causes intra-phase conflicts between bus and car lanes (shown in Figure 5.4). A conflict avoidance technique to eliminate these intra-phase conflicts will be discussed later. The signal phasing convention is formulated using the following equality constraints.

$$\Omega^t_{(z,p)_\ell} = \Omega^t_{(z,p+4)_\ell},\ z \in \mathcal{I},\ p \in \{1, 3\},\ (z, p) \in \mathcal{C}_I,\ \ell \in \{B, G\},\ \forall t \in \mathcal{T}. \tag{5.39}$$

$$\Omega^t_{(z,p)_\ell} = \Omega^t_{(z,p+4)_\ell} = \Omega^t_{(z,p+7)_\ell} = \Omega^t_{(z,p+8)_\ell},\ z \in \mathcal{I},\ p \in \{2, 4\}. \tag{5.40}$$

where, $(z, p) \in \mathcal{C}_I,\ \ell \in \{B, G\},\ \forall t \in \mathcal{T}$. Equation (5.39) allows right-turn traffic movements in phase 1 and 3. Whereas, Equation (5.40) synchronizes through and left movements in phase 2 and 4. The following constraint ensures that both the bus and car classes in the same phase get the same green-time.

$$\Omega^t_{(z,p)_B} = \Omega^t_{(z,p)_G},\ z \in \mathcal{I},\ p \in \{1, 2, 3, 4\},\ (z, p) \in \mathcal{C}_I,\ \forall t \in \mathcal{T}. \tag{5.41}$$

The minimum and maximum phase lengths are defined by the following constraints so that a fair-distribution of the green-time among the phases can be assured when required.

$$\Omega^t_{(z,p)_\ell} \geqslant \Omega^{min}_{(z,p)_\ell},\ z \in \mathcal{I}, P = \{1, 2, 3, 4\},\ (z, p) \in \mathcal{C}_I,\ \ell \in \{B, G\},\ \forall t \in \mathcal{T}. \tag{5.42}$$

$$\Omega^t_{(z,p)_\ell} \leqslant \Omega^{max}_{(z,p)_\ell},\ z \in \mathcal{I}, P = \{1, 2, 3, 4\},\ (z, p) \in \mathcal{C}_I,\ \ell \in \{B, G\},\ \forall t \in \mathcal{T}. \tag{5.43}$$

In Equation (5.42) and (5.43), $\Omega^{min}_{(z,p)_\ell}$ and $\Omega^{max}_{(z,p)_\ell}$ are predetermined minimum and maximum green-times, respectively. Finally, one must ensure that the sum of the green fractions does not exceed 1, this leads to the constraint

$$\sum_{\forall p \in P} \Omega^t_{(z,p)_\ell} \leqslant 1,\ z \in \mathcal{I}, P = \{1,2,3,4\},\ (z,p) \in \mathcal{C}_I,\ \ell \in \{B,G\},\ \forall t \in \mathcal{T}. \tag{5.44}$$

Intra-phase conflicts occur between the bus and car cells in phase 2 and 4 when the through and the left movements are allowed together. The intra-phase conflicting movements for the mentioned phases are shown in Figure 5.4. To avoid this conflict, we introduce the following two constraints.

$$y^t_{(z,p+q)_G j_G} \leqslant \omega^t_{(z,p+q)_G} Q^t_{(z,p+q)_G},\ p \in \{2,4\},\ q \in \{7,8\},\ (z,p+q) \in \mathcal{C}_I. \tag{5.45}$$

$$y^t_{(z,p+q)_B j_B} \leqslant \omega^t_{(z,p+q)_B} Q^t_{(z,p+q)_B},\ z \in \mathcal{I},\ p \in \{2,4\},\ q \in \{0,4\},\ (z,p) \in \mathcal{C}_I. \tag{5.46}$$

In the above constraints $z \in \mathcal{I},\ t \in \mathcal{T}$. The conflict avoidance variable $\omega^t_{(z,p+q)_G}$ in Equation (5.45) controls the outgoing flow from the intersection car cell $(z,p+q)_G$ (e.g., 109, 111, 110, 112 in Figure 5.4) for the phases $p \in \{2,4\}$. Similarly, the decision variable $\omega^t_{(z,p+q)_B}$ in Equation (5.46) is used to introduce control on flow out of the intersection bus cell $(z,p+q)_B$ (e.g., 102, 104, 106, 108 in Figure 5.4). Finally, the sum of these disaggregated green-splits must not exceed the phase green-split $\Omega^t_{(z,p+q)}$. Constraints (5.47)-(5.48) ensure that the sum of the disaggregated green-splits for the conflicting bus cell (i.e., $(z,p)_B$) and car cell (e.g., $(z,p+7)_G$) does not exceed the total allocated green-split to the corresponding phase.

$$\omega^t_{(z,p)_B} + \omega^t_{(z,p+7)_G} \leqslant \Omega^t_{(z,p)},\ z \in \mathcal{I},\ p \in \{2,4\},\ (z,p) \in \mathcal{C}_I,\ t \in \mathcal{T}. \tag{5.47}$$

$$\omega^t_{(z,p+4)_B} + \omega^t_{(z,p+8)_G} \leqslant \Omega^t_{(z,p)},\ z \in \mathcal{I},\ p \in \{2,4\},\ (z,p) \in \mathcal{C}_I,\ t \in \mathcal{T}. \tag{5.48}$$

For example, due to constraint (5.47) the sum of the allocated green-splits to the bus cell 102 and car cell 109 would not exceed phase-2 green-proportion Ω^t_{102} at time-slot t. The complete BP SO-DTA-SC framework is presented in the following sub-section.

5.2.9 The Complete BP SO-DTA-SC Formulation

BP SO-DTA-SC :

minimize $\quad \mathcal{O} =$ Equation (5.1)

subject to $\quad$ BP SO-DTA constraints (5.2)-(5.32) and

TPE-SC constraints (5.33)-(5.48).

$$\sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C}_B \backslash \mathcal{C}_{S_B}} \alpha_B \cdot x^t_{i_B} + \sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C}_G \backslash \mathcal{C}_{S_G}} \alpha_G \cdot x^t_{i_G} = \mathcal{O}^\star_{TSPT}. \quad (5.49)$$

When $\mathcal{O}^\star_{TSPT}$ is attained by solving the following problem

$$\mathcal{O}_{TSPT} = \min \sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C}_B \backslash \mathcal{C}_{S_B}} \alpha_B \cdot x^t_{i_B} + \sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C}_G \backslash \mathcal{C}_{S_G}} \alpha_G \cdot x^t_{i_G}, \; \alpha_B > \alpha_G, \quad (5.50)$$

subject to the constraints (5.2)-(5.32), (5.33)-(5.48).

The BP SO-DTA-SC formulation presented above combines DBLs for the bus vehicles and the TPE-SC model that can handle TSP at the intersections. Furthermore, one can assign routes to the buses by using this formulation. The objective function ensures that the solution is NHB. The bus vehicles are given priority over cars till total number of occupants in the waiting cars exceeds that of the bus (see Remark 5). The Equation (5.50) was presented by Liu et al. (2015a). In this equation $\alpha_B > \alpha_G$, as a result, this objective gives priority to the buses if there is conflict between bus and car lane traffic movements. Although equation (5.50) cannot resolve the HB problem, this attains SO TSPT. As a result, the inclusion of the constraint (5.49) in our formulation ensures that the formulation attains SO solution (i.e., passenger throughput optimal) while remaining NHB.

Remark 5. *Beside giving priority to the buses, the proposed BP SO-DTA-SC formulation also assures fairness to the car users at the conflicting (between car and bus movements) diverging cells, merging cells, and at the intersections. If a bus cell* i_B

and a car cell i_G *is located in the network such that they cross each other's path and* $\Psi_{i_B} = \Psi_{i_G}$*, then in the objective function* (5.1)*, when* $\alpha_B \times x^t_{i_B} < \alpha_G \times x^t_{i_G}$*, the cars get the opportunity over bus vehicles to move forward. For example, if we assume* $\alpha_B = 15$*,* $x^t_{i_B} = 1$*,* $\alpha_G = 1.5$*, and* $x^t_{i_G} = 20$*, then* $15 \times 1 < 1.5 \times 20$ *and cars would get the opportunity to move forward. As a result, while remaining NHB, this property of the proposed method assures that the cars do not face excessive congestions.* $\odot$

In the Proposition B.2.1 outlined in B.2 we show that this formulation can indeed resolve the HB problem. One can measure Total System-wide Travel Time (TSTT) (for the vehicles, not for the passengers) from the optimal solution of the BP SO-DTA-SC formulation using the equation provided below. The definition of TSTT was proposed by Ziliaskopoulos (2000) for a single-class traffic setting. Here we apply that to a multi-class traffic scenario.

$$\text{TSTT} = \sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C}_B \setminus \mathcal{C}_{S_B}} x^{\star t}_{i_B} \cdot \tau + \sum_{\forall t \in \mathcal{T}} \sum_{\forall i \in \mathcal{C}_G \setminus \mathcal{C}_{S_G}} x^{\star t}_{i_G} \cdot \tau. \tag{5.51}$$

In Equation (5.51), $x^{\star t}_{i_B}$ is optimal bus occupancy, $x^{\star t}_{i_G}$ is optimal car occupancy, and τ is the time-slot duration. Note that the TSTT does not include the passenger multiplication factor like in the TSPT calculation in Equation (5.50).

5.3 Results and Discussions

In this section results related to the presented BP SO-DTA and TPE-SC models are outlined. The objectives of these results are to investigate whether

- the model can implement bus priority and what are the impacts of bus priority on the optimal solution;
- the TSPT can be minimized by introducing DBLs on the roads;
- the total intersection delay of the buses is also minimized (i.e., TSP); and

- the presented BP SO-DTA-SC formulation can be applied to analyze city-size networks.

5.3.1 Performance of the BP SO-DTA-SC Formulation

The BP SO-DTA-SC formulation presented in Section 5.2.9 is free of vehicle HB problem. Furthermore, the BP SO-DTA-SC provides priority to the buses on the links and at the intersections. The mixed-integer formulations presented in the literature allocate the total duration of a time-slot to the bus vehicle class when a conflicting-traffic-movement between bus and car is detected. However, in many instances, the allocated time-slot to a particular group of bus vehicles would not be fully utilized which would impose an unnecessary delay on the cars. In contrast, in the proposed BP SO-DTA-SC method, only a portion of the time-slot is allocated to the buses. This portion is as much as required by the bus vehicles to move forward. Furthermore, in our proposed formulation the priority is given according to the number of passengers waiting in the buses or cars to be served. As a result, the proposed approach considers a fair distribution of delay to the cars. This section analyses performance of the BP SO-DTA-SC framework regarding TSPT, TSTT, and priority.

To attain an optimal solution of the combined traffic assignment and signal control problem, the 60 seconds (time-slot duration) CTM network presented in Figure 5.4 is implemented. This network consists of four source cells, one diverging cell, several ordinary cells, a few merging cells, and two sink cells belonging to each of the bus and car vehicle class. The conflicting movements between bus merging cells (diverging cell) and car merging cells (diverging cell) are also visible in this figure. Intra-phase conflicting movements are also shown.

The physical parameters of the network are presented in Table 5.1. In the table, observe that the cells 1, 2, 7, 17, 18, 19, 20, 21, 22 have higher flow and occupancy capacity than the other cells so that a greater number of vehicles can enter and exit the network simultaneously. The merging cells (e.g., 2, 7) at the intersection have higher capacity than the intersection cells (e.g., 101, 102). This is to make sure that these merging cells are not imposing bottlenecks, therefore, help to attain increased flow at the intersection. Both the bus and car cells are assumed to have the same free-flow speed and backward propagation speed. As a result, they have the same cell-length. Note that the BP SO-DTA-SC supports different cell-lengths for the bus and car vehicle class. The input demands in all the bus source cells are $\{5, 10, 15, 20, 30\}$ bus vehicles at the first five consecutive time-slots. Similarly, $\{30, 40, 50, 60, 70\}$ cars are generated at the car source cells during the first five time-slots. These varying car and bus demands are used to create different traffic conditions (e.g., under-saturated, over-saturated). The TSP-SC model is also implemented at the intersection. For simplicity, we ignored the phase loss, minimum green-time, and maximum green-time constraints.

In Table 5.1, the aggregated bus and car network parameters in case of the HB SO-DTA are outlined. The bus demands are converted to equivalent car demands. Then the sum of these equal car demands is added to the original car demands and used as inputs to the source cells in the HB SO-DTA scenario. For example, in the first time-slot, the bus demand is 5, and car demand is 30. If we consider 30 passengers per bus and 1.5 passengers per car, then one bus demand is equivalent to 20 cars. As a consequence, total input demand in HB SO-DTA in the first time-slot is 130 cars. Note that throughout this chapter we assume an average of 1.5 passengers per car. Whereas, in the case of bus vehicle, different numbers of passengers (e.g., 30, 15, 9 passengers per bus) are considered. The optimization time horizon was set to 100 time-slots (i.e., 100 minutes). In the BP SO-DTA-SC case, bus routes are set using Equation (5.32) in such a way that at least 60% (i.e., $0.6 \times 320 = 192$ buses) of the total bus demand go through bus cell 15 and the rest 40% (i.e., $0.4 \times 320 = 128$ buses) of the total demand go through bus cell 16. Furthermore, to ensure that the longest

Table 5.1. The physical parameters of the network.

Network	Parameters				
	Cell index	Q_i^*	N_i^*	v_i (ms^{-1})	δ_i
bus cells	1, 2, 7, 17, 18, 19, 20, 21, 22	36	108	15.24	0.5
	R1, R2, R3, R4	36	∞	-	-
	all others	12	36	15.24	0.5
car cells	1, 2, 7, 17, 18, 19, 20, 21, 22	108	324	15.24	0.5
	R1, R2, R3, R4	108	∞	-	-
	all others	36	108	15.24	0.5
HB SO-DTA cells†	1, 2, 7, 17, 18, 19, 20, 21, 22	144	432	15.24	0.5
	R1, R2, R3, R4	144	∞	-	-
	all others	48	144	15.24	0.5

†equivalent parameters of the network for the HB SO-DTA formulation; *veh per slot

route (through bus cell 9) has been used, a minimum of 10% of the bus vehicles are allocated to this route. This is just an example scenario to show how one can implement bus routes using the proposed formulation. In the HB SO-DTA case, there are no predefined routes. Three scenarios are studied for both the BP SO-DTA-SC and HB SO-DTA formulations. The scenarios are as follows:

1. $\alpha_B = 9$ and $\alpha_G = 1.5$;
2. $\alpha_B = 15$ and $\alpha_G = 1.5$; and
3. $\alpha_B = 30$ and $\alpha_G = 1.5$.

The optimal solutions attained for these three scenarios are discussed in the following sections.

5.3.1.1 Impact of α_B on Bus and Car Travel-time

In Figure 5.5, bus TSTT, car TSTT, and TSTT for all the vehicles attained using the BP SO-DTA-SC formulation are presented for the three scenarios discussed above. The figure depicts that with increasing bus priority (i.e., the higher value of α_B), the travel-time for the buses is reduced while cars face higher delay due to their less

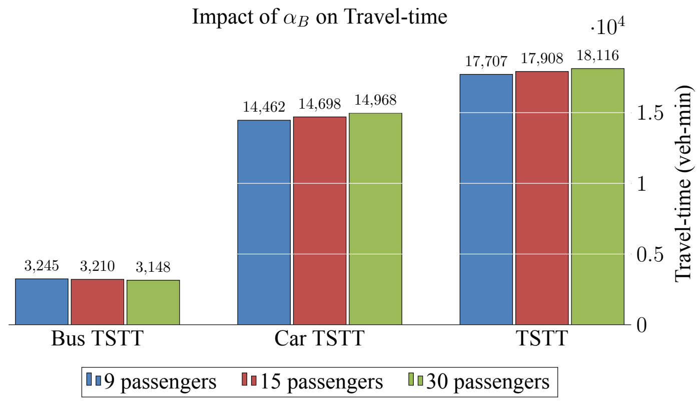


Figure 5.5. The TSTTs attained by the BP SO-DTA-SC formulation for different scenarios are shown in the figure.

priority in the network. This is expected as for the higher value of α_B (we kept $\alpha_G = 1.5$), the cars are delayed till a total number of passengers in the waiting cars exceeded that of the buses. As a result, the priority is given in terms number of people waiting to be served either in the cars or the buses. Therefore, the fairness to the car occupants is also assured so that car-passengers would not suffer excessive delay. This property of the BP SO-DTA-SC model is also clarified in Remark 5. The difference between the car TSTTs when $\alpha_B = 9$ and $\alpha_B = 15$ is only 1.6%. Whereas, for the $\alpha_B = 9$ and $\alpha_B = 30$ scenarios, the difference in car travel time is 2.3%. This result verifies the fact that once the bus priority is already given (i.e., $\alpha_B > \alpha_G$), further increase in α_B has a minor impact on the TSTT that would be attained by the BP SO-DTA-SC model. The TSTT may increase by a marginal amount, but this DBLP system would facilitate to increase the passenger throughput which we will see in the next section.

5.3.1.2 Performance and Complexity Associated to the DBLP

The TSPT obtained using the proposed BP SO-DTA-SC and HB SO-DTA in Ziliaskopoulos (2000) are presented in Table 5.2. In case of HB SO-DTA, no traffic signal control and conflict avoidance constraints are included. For the tested three scenarios, the HB SO-DTA attains TSTT of 52800 veh-min, 103784 veh-min, and

Table 5.2. Comparison of passenger travel-time.

Scenarios	TSPT BP SO-DTA-SC	TSPT HB SO-DTA‡	Improvement
1†	50898	79200	56%
2†	70197	155676	122%
3†	116892	458868	293%

‡conflict avoidance and signal control constraints are not considered; †all the bus and car demands are converted into equivalent number of passengers.

305912 veh-min, respectively. In the Table 5.2, it is noticeable that the proposed unified BRT-TSP framework (i.e., BP SO-DTA-SC model) significantly reduces (e.g., by 56% in scenario-1) the passenger travel-time over the HB SO-DTA formulation. The results also show that with the increasing number of passengers per bus, the BP SO-DTA-SC passenger travel time is reduced substantially. For instance, in the case of 30 passengers occupying a bus, the improvement in TSPT is 293% compared to the HB SO-DTA. This result verifies that by introducing dedicated lanes and priority for the buses on the diverging and merging links and at the intersections, the traffic congestion as well as the passenger travel-time can be significantly reduced.

The associated computational complexity (i.e., the sum of the total number of active variables and constraints) for the three scenarios discussed above is presented in Table 5.3. For scenario-1, compared to the HB SO-DTA formulation, the BP SO-DTA-SC model has higher complexity due to the increased number of cells per road segments (i.e., bus and car cells for the respective lanes), the traffic signal control constraints, and the conflict avoidance variables at the diverging or merging cells. Nevertheless, due to the linear formulation, the BP SO-DTA-SC formulation took maximum 3s of time to reach the optimal solutions. The complexity of the HB SO-DTA formulation increases (i.e., scenario-3) and requires more time to solve when the number of passengers per bus increases. This is because the total demand increased with increasing number of passengers and required longer optimization time horizon

Table 5.3. Comparison of complexity and solution time.

	BP SO-DTA-SC			HB SO-DTA		
Scenarios	Network cleared	Time to solve	CX*	Network cleared	Time to solve	CX*
1†	29 (mins)	2s	18255	38 (mins)	1.5s	9392
2†	29 (mins)	2s	18255	51 (mins)	2s	12850
3†	30 (mins)	3s	18885	85 (mins)	4.5s	21248

†all the vehicles were converted into equivalent number of passengers; *complexity (CX) is the sum of active variables and constraints.

to clear the network. In the table we see that the HB SO-DTA required 38, 51, and 85 minutes to clear the network for the scenarios 1, 2, and 3 respectively while BP SO-DTA-SC always took less or equal to 30 minutes. As a result, the TSTT for the HB SO-DTA case is significantly higher, that eventually increases the TSPT.

5.3.1.3 Bus-priority vs No-bus-priority in BP SO-DTA-SC

In this section, we investigate two cases: (i) priority is given to the buses, and (ii) no priority is given to them in the proposed BP SO-DTA-SC formulation. To see how they differ, we compare TSTT and intersection delay of these two scenarios for the network shown in Figure 5.4. The scenarios are described below:

(I). the BP SO-DTA-SC formulation with $\alpha_B > \alpha_G$ in the objective function (5.1);

(II). the BP SO-DTA-SC formulation with $\alpha_B = \alpha_G$ in the objective function (5.1). In this scenario, the objective function would not provide any priority to the buses.

The resulting TSTT and TSPT for the two scenarios are presented in Table 5.4. In the table, we observe that giving priority to the buses improved the TSPT by 26% but degraded TSTT by 16%. This increase in TSTT is expected as cars had to wait so that overall TSPT could improve. The result also shows that the total intersection delay experienced by the buses was significantly reduced by 78.5%. This result validates the fact that the presented signal control model effectively performed

Table 5.4. Impact of bus priority on TSTT, intersection delay, and TSPT.

	Performance					
Scenarios	TSTT	Deterioration	IBD††	Improvement	TSPT*	Improvement
Scenario (I)†	17908	16%	387.3	78.5%	70197	26%
Scenario (II)‡	15029.3	-	691.19	-	88063	-

††Intersection Bus Delay (IBD); †using the BP SO-DTA-SC formulation without priority; ‡using the BP SO-DTA-SC formulation with priority; *by setting 1 bus = 15 passengers and 1 car = 1.5 passengers.

to implement TSP when priority was actually given through the objective function in Equation (5.1).

5.3.1.4 Performance of the Proposed TPE-SC Model

To assess the performance of the TPE-SC model, we consider following two cases and solve the network in Fig. 5.4 with input demands and physical parameters presented in Sec. 5.3.1. These cases are described below:

TPE-SC case: the BP SODTA-SC formulation with $\alpha_B = 15$ and $\alpha_G = 1.5$ in the objective function (5.1);

MTSP case: the BP SODTA formulation with Mixed-integer Transit Signal Priority (MTSP) formulation presented in B.3 when all other settings kept the same as the TPE-SC case.

The results for these two cases are presented in Tab. 5.5. Observe that the TPE-SC case attained improved TSTT (15.6%), TSPT (9%), and intersection car delay (40%). This is expected as the MTSP control method is binary in nature and only can allocate the whole time unit (i.e., a 60s cycle in this example) to a phase. Moreover, cars always needed to wait for green signal if a bus arrives at the intersection which caused excessive delay to the cars (i.e., no fairness). Despite giving priority to bus, the MTSP attained lower performance (-15.6%) in terms of intersection bus delay. It is because of the competing demands between inter-phase bus movements worsening by the mixed-integer nature of the model. In Tab. 8 we also show that the TPE-SC model has lower

Table 5.5. Comparison between TPE-SC and MTSP control methods.

	TPE-SC case	MTSP case	Improvement††
TSTT	17908 (veh-min)	20712 (veh-min)	15.6%
TSPT	70197 (passenger-min)	76293 (passenger-min)	9%
Intersection car delay	2944 (veh-min)	4128 (veh-min)	40%
Intersection bus delay	387.3 (veh-min)	476 (veh-min)	23%
Complexity*	18255	20415	12%
Solution time	2s	680s	33900%

††improvement with respect to the TPE-SC case; *sum of active variables and constraints.

complexity and the optimal solution converged much faster than that of the MTSP model. For further demonstration of this property of the BP SODTA formulation and TPE-SC model, we analyze a city-size network in the following section.

5.3.2 Scalability of the BP SO-DTA-SC Formulation

The proposed BP SO-DTA-SC formulation is a linear framework and computationally tractable. As a result, this model can be applied to analyze realistically large networks. To show scalability of our formulation, we analyze the Fort-Worth network (Mahmassani 2001) presented in Figure 5.6. This network is a city-wide network in Texas and spread over thirteen zones. The network consists of 441 links and 168 nodes. We consider a node in the arterial streets as an intersection when four links intersect at that node. As a result, there are total 62 intersections in the network. The two interstate freeway links, having the length of 50km, are connected to the two 50 km supporting lanes. The free-flow speed of the freeways is 100km/hr. We also assume all other links have a length of 6km and speed limit of 60km/hr. Based on these free-flow speeds, the corresponding links were converted into cells. The inputs to the BP SO-DTA-SC model is presented in Table 5.6. Optimization was run in a 64-bit Windows 7 desktop computer consisting of 2.5 GHz processor and 8GB RAM.

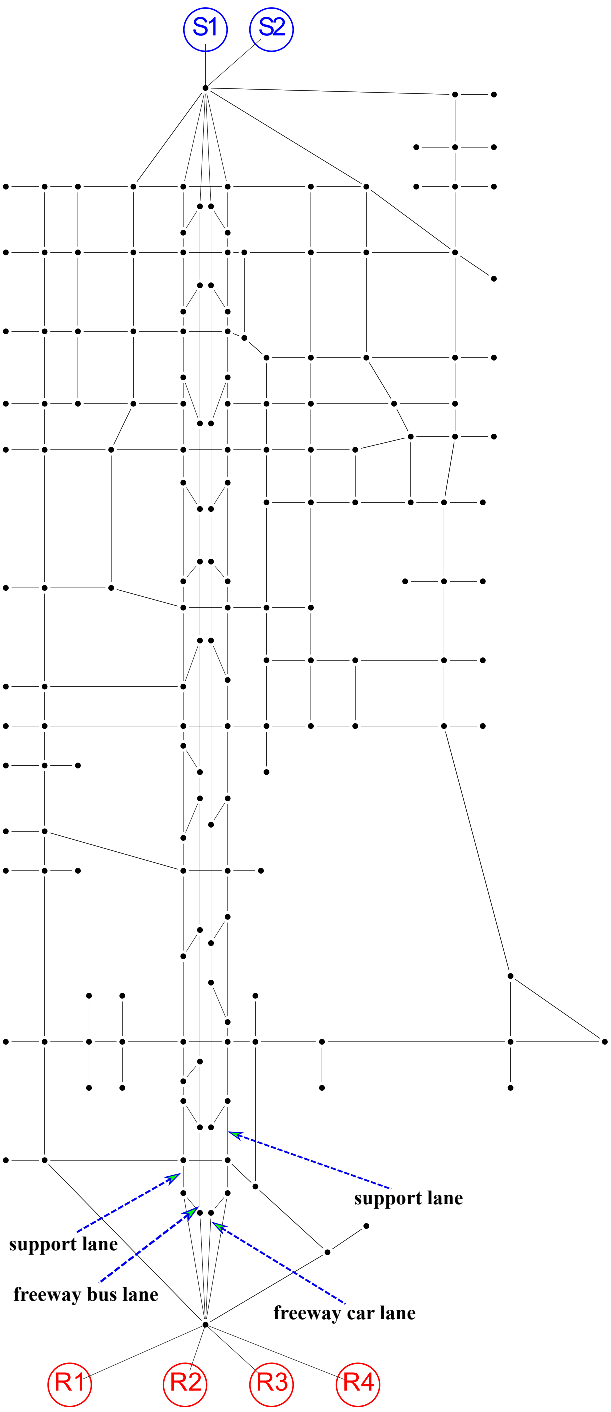


Figure 5.6. The Fort-Worth network under investigation is shown in the figure. There are four vertical links in the middle. Each of these links has a length of 50 kilometers. Out of these four links, we consider two inner links as freeway lanes, where the left freeway lane is a BRT lane, and the right freeway lane is a car lane. The two outer links of the bus freeway and car freeway lanes are supporting links to enter and exit the freeways. R1 and R2 are bus source nodes, and S1 is the bus sink node. Whereas, R3 and R4 are car source nodes and S2 is the car sink node.

Table 5.6. Physical parameters, demands, and settings for the Fort-Worth network.

All the freeway cells	Q_i (veh per slot)	N_i (veh per slot)	v_i (km/hr)	δ_i	$L^\dagger$
	33	99	100	0.5	833m
All other cells	Q_i (veh per slot)	N_i (veh per slot)	v_i (km/hr)	δ_i	$L^\dagger$
	20	60	60	0.5	500m
Time-slot duration	30s				
Bus demands at $R1$ and $R2$	$\{5,5,5,5,5,5,5,5,5,5\}^\ddagger$ (total 100 buses)				
Car demands at $R3$ and $R4$	$\{10,20,30,40,50,60,70,80,90,100\}^\ddagger$ (total 1100 cars)				
Optimization time-horizon	120 time-slots or 1hr				

‡arriving at first ten consecutive time-slots; †cell-length in meters (m) for both the bus and car cells.

Three scenarios were tested using the network settings presented in the Table 5.6. The scenarios are described below.

1. Equation (5.50) as objective, subject to the constraints (5.2)-(5.32), (5.33)-(5.48).

2. The BP SO-DTA-SC formulation.

3. Equation (5.51) as objective with constraints (5.2)-(5.32), (5.33)-(5.48).

Results for these three formulations are presented in Table 5.7. In the table we see that the total number of active variables and constraints (i.e., complexity) that was used by these three formulations was approximately 0.9 millions (mil). Given that high complexity, it took maximum of 189 seconds (i.e., for scenario 1) to attain optimal solutions for this city-size network. This result verifies the fact that the proposed linear BP SO-DTA formulation and the linear TPE-SC model are computationally tractable and can be applied to analyze large scale BRT-TSP system.

In the table, also notice that the bus travel-time improved significantly by 107% in scenario 1 and 2 when priority was given to them. However, this bus-priority increased car travel-time by a minor 5%. Similar result was found in the previous section. Compared to the non-bus-priority scenario 1, the overall improvement in TSPT in the two bus priority scenarios was 28%. This result was obtained when we considered 15 occupants per bus. Therefore, it would have further improved if we had considered more occupants per bus. The key outcomes of these results are that although the proposed BRT-TSP framework may impose minor unwanted delay to the car users, it significantly improves throughput of the network. Furthermore, this framework is computationally tractable and can be used to analyze city-size DTA problems in a car-bus scenario.

Table 5.7. Performance of BP SO-DTA-SC when analyzing the Fort-Worth network.

	Performance					
Scenarios	Complexity	Solution time	Bus TSTT	Car TSTT	TSPT*	Improvement††
1	≈0.9 (mil)	189s	24.1hr	584hr	1237.5hr	28%
2	≈0.9 (mil)	186s	24.1hr	584hr	1237.5hr	28%
3	≈0.9 (mil)	185s	50hr	554hr	1566hr	-

††improvement in TSPT with respect to scenario 3; *by setting 1 bus = 15 passengers ($\alpha_B = 15$) and 1 car = 1.5 passengers ($\alpha_G = 1.5$).

5.4 Chapter Summary

In this chapter, we formulated the BRT-TSP system as a *linear BP SO-DTA-SC* framework. This formulation supports multi-class traffic for single destination networks, it can model bus routes and attains SO solutions in a bus-priority setting. Unlike other formulations in the relevant field, this formulation attains NHB solutions and gives priority to the buses by assuring fairness to the all the passengers. Furthermore, the embedded signal control model in the framework namely TPE-SC resolves inter-phase and intra-phase conflicts between bus and car movements at the intersections.

We found that the TSPT was significantly reduced (up to 293%) by introducing separate lanes for the buses. Although the BP SO-DTA-SC approach has a higher number of decision variables and constraints, due to the linear formulation, it attained the solution within a very short time. Even for the large Fort-Worth network, the proposed model reached optimal solutions within 3.5 minutes. As a result, this model is computationally tractable and can be applied to analyze city-size BRT-TSP systems.

We also found that there is a trade-off between TSPT or intersection delay and TSTT when priority to the buses is considered. Numerical solutions showed that when priority was given to the bus vehicles, then the average TSTT for all the vehicles was increased by 16%. However, the passenger travel-time and intersection delay decreased by 26% (28% in the Fort-Worth scenario) and 78.5%, respectively.

These results indicate that DBLP may increase average vehicle travel-time, by a minor amount, but the overall passenger throughput is significantly improved. This would motivate a higher number of people to use the public transport.

Chapter 6

Conclusion and Future Work

Dynamic Traffic Assignment (DTA) is a mathematical formulation that provides a basis for the optimization of vehicular traffic network performance and planning. These types of formulations are currently popular modeling tools for tackling various applications such as transportation system planning and designing traffic controllers at intersections. In this thesis, we have explored several aspects of Non-Holding-Back (NHB) System Optimal (SO) traffic assignment, and Signal Control (SC). In the following sections the major contributions of this thesis, key results, and future research directions are outlined.

6.1 The NHB DTA-SC Formulation

In Chapter 3, we developed a novel *linear and continuous NHB DTA-SC* framework. The proposed framework attains SO solutions for a combined signal control and dynamic traffic assignment problem and resolves the known vehicle Holding-Back (HB) problem. Our contributions related to the NHB DTA-SC formulation are,

1. the inclusion of an objective function which eliminates the vehicle HB problem;
2. we proposed the *Signal Control with Realistic Cycle (SCRC)* model which eliminates the trade-off between accuracy and complexity (i.e., cycle-length vs. cell-length).

We have also found that our proposed SCRC model

- achieves travel times comparable to that of the existing CSDT control scheme,
- and for the same CTM time-scale, it has lower computational complexity.

Moreover, the SCRC model is more adaptive to varying traffic conditions, thus, making it more effective in reducing congestion compared to the popular MISC model. The SCRC model enables us to use a realistic cycle-length as well as a shorter discrete time-slot for higher accuracy of the DTA solution. The SCRC also induces fewer decision variables compared to the other models and reaches the optimal solution faster.

The accuracy of a DTA solution depends on the time-slot duration of the DTA problem. The shorter the time-slot duration, the more accurate the DTA solution becomes. While analyzing the impact of time-slot duration on the SO solution, we have found that the traffic density in the DTA solution significantly changes with varying time-slot duration. As an example, for the tested 60s and 5s time-slots, there is 17.5% difference in TSTT between their solutions. As traffic density determines vehicle speed and level of congestion, the time-slot duration also influences the accuracy of emission estimation in a DTA framework. This fact motivated us to add an emission perspective in the DTA formulation and study the impact of time-slot duration on the emission based DTA solution. The contributions of this thesis related to the emission topic are given below.

1. We have proposed a linear *NHB Emission-Based DTA with Signal Control (EB DTA-SC) formulation*. To the best of our knowledge, this is the first *linear DTA-SC formulation with emission perspective*.
2. We have also proposed a framework to estimate emissions at the post-processing stage using a DTA solution.

While analyzing emission solutions, we found that there was up to 32% difference between the NO_X emission measurements using the 5s and the 60s time-slots. This result also verifies that the DTA time-scale has a significant impact on the accuracy of the DTA solution. In the next section, we summarize the properties of the optimal solutions.

6.2 Structures of the Optimal Solutions

In Chapter 4, we have presented a number of optimal solution structures of an NHB DTA-SC formulation. By primal analysis of this formulation, we have affirmed and mathematically proven the following properties of the optimal traffic assignment process.

1. In general, an NHB DTA-SC solution is not unique, and the formulation can attain multiple solutions by interchanging occupancies among the cells within the same penalty levels. However, for a single path network, it obtains a unique solution.

2. The NHB DTA-SC formulation maximizes traffic flow throughout the whole network. In other words, the formulation minimizes network-wide delay as well as the delay at the intersections.

In the second part of Chapter 4, we derived the structural properties of the optimal signal control policy in the proposed NHB DTA-SC formulation by Lagrangian dual decomposition and constraint qualifications. Our core findings related to the optimal signal control settings are:

1. in the case of an under-saturated traffic condition, the optimal green-time is proportional to the available demand at the corresponding intersection cell;

2. optimal control follows Webster's "equisaturated" (Webster 1958) control in an over-saturated traffic state; and

3. in a queue spill-back scenario, the green-times are determined by available spaces at the downstream links.

We have also mathematically proven these control structures. The numerical solutions verified them by showing exact convergence. These unique signal control structures are yet to be presented in the literature and novel contributions in the field. These structures are locally optimal and ready to implement ("off the shelf") in the intersection traffic controllers. In the next section, we summarize the findings related to the *Bus Rapid Transit with Transit Signal Priority (BRT-TSP)* contents proposed in this thesis.

6.3 A Novel DTA Model for BRT-TSP Systems

One of the primary objectives of this research was to develop a realistic and computationally tractable linear programming formulation that can accommodate priority vehicles and non-priority general traffics within the same framework. In Chapter 5, we proposed a novel linear programming formulation namely *Bus Priority System Optimal Dynamic Traffic Assignment (BP SO-DTA)* model, that can support a multi-class traffic environment. We have also introduced a linear *Transit Priority Enabling Signal Control (TPE-SC)* model. The combined traffic assignment and signal control framework is named as *Bus Priority System Optimal Dynamic Traffic Assignment with Signal Control (BP SO-DTA-SC)* formulation. This computationally tractable unified BRT-TSP framework is a significant contribution in the field of DTA. The key features of the proposed framework are:

- It is a linear and NHB formulation.
- It enables us to model priority give-way rules.
- It can model bus routes.

- The embedded TPE-SC model resolves intra-phase conflicts.
- The formulation assures fairness to the all the road users.
- It can be applied to analyze city-size networks.

In Chapter 5 we also studied the impact of introducing the *Dedicated Bus-only Lanes with Priority (DBLP)* on the *Total System-wide Passenger Travel-time (TSPT)*. We have found that DBLP can reduce the TSPT up to 293% when compared to the existing non-lane based HB case.

We also investigated the impact of bus priority on the passenger travel-time (i.e., TSPT) and vehicle travel-time (i.e., Total System-wide Travel Time (TSTT)). We compared a priority lane-based case to a non-priority lane-based case. The results showed that when bus priority was implemented then

- the TSPT decreased by 26%,
- intersection delay reduced by 78.5%, and
- TSTT increased by 16%.

These results suggest that the BRT-TSP system may increase average vehicle travel-time, by a minor amount (e.g., 16%), but the overall passenger throughput would significantly improve (e.g., 293%). This would motivate a higher number of people to use the public transport over the private cars, which will reduce congestion and vehicle discharged emissions further. The next section discusses interesting research problems that can be explored further.

6.4 Future Work

This thesis has revealed several research directions that are discussed below.

- Firstly, the linear NHB DTA-SC formulation presented in this thesis can be applied to analyze large traffic networks and would enable one to detect bottleneck links. This approach would further help to determine to what extent the capacity of the bottleneck should be increased. Increasing link capacity requires funding and the acquisition of the necessary land. If a DTA problem involves such budgetary constraints, then the problem is called *Network Design Problem (NDP)*. The investigation of the NDP in a linear NHB context is an area that could be considered for future research.

- Secondly, we have presented a marginal cost analysis for a decomposed problem of the proposed NHB DTA-SC formulation. Similar dual decomposition method can be carried out for the User Equilibrium Dynamic Traffic Assignment with Signal Control (UE DTA-SC) problem. Under certain traffic conditions, the UE DTA-SC becomes SO. As a result, it might be interesting to analyze circumstances for which the UE solution becomes SO and possible traffic states for which the UE and the SO solutions differ.

- In an HB SO-DTA solution, surplus demands are held back at the origins, and the network runs at the saturation flow capacity of the bottleneck. By doing so, all the vehicles can travel at the free-flow speed. One of the potential applications of this property is ramp-metering at the freeway entrances and associated reservoir-link planning. Scrutinizing this problem under a stochastic demand scenario might be interesting.

- Finally, emission perspective can be added to the BRT-TSP system. In this thesis, the benefits of this system are shown concerning TSPT. It would be interesting to see how the emission objective influences the travel-time, associated bus-car interaction, and whether it jointly minimizes emission, TSPT, and TSTT.

Appendices

Appendix A

A.1 Generating penalty terms: An example

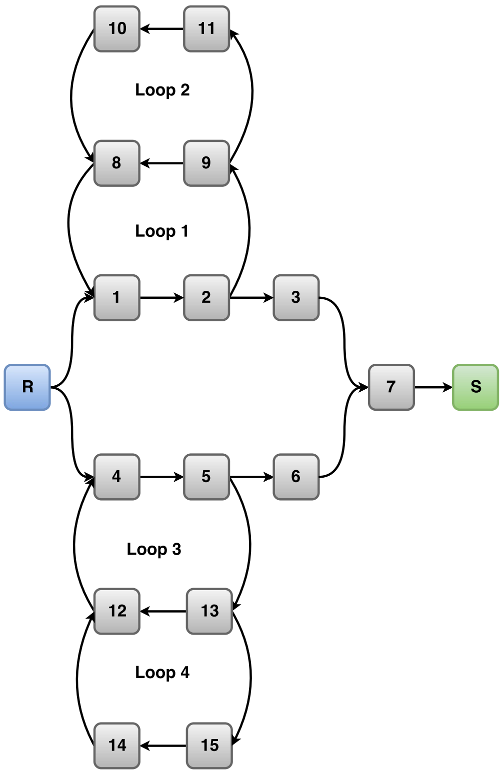


Figure A.1. The example network for which Ψ_i values are determined. There are four cyclic loops in the network. Loop 1 and 2 are symmetric to loop 3 and 4 with respect to the source and sink cell.

The algorithm was presented in Section 3.2.2.1. For the network in Figure A.1, we outline the output of the algorithm below.

Start

1. Initialization:

(a) $\Psi_S = 1$.

(b) Determining set of all the paths $\mathcal{P}$:

i. Path-1 (P1): $S \rightarrow 7 \rightarrow 3 \rightarrow 2 \rightarrow 1 \rightarrow R$

ii. Path-2 (P2): $S \rightarrow 7 \rightarrow 3 \rightarrow 2 \rightarrow 1 \rightarrow 8 \rightarrow 9$

iii. Path-3 (P3): $S \rightarrow 7 \rightarrow 3 \rightarrow 2 \rightarrow 1 \rightarrow 8 \rightarrow 10 \rightarrow 11 \rightarrow 9$

iv. Path-4 (P4): $S \rightarrow 7 \rightarrow 6 \rightarrow 5 \rightarrow 4 \rightarrow R$

v. Path-5 (P5): $S \rightarrow 7 \rightarrow 6 \rightarrow 5 \rightarrow 4 \rightarrow 12 \rightarrow 13$

vi. Path-6 (P6): $S \rightarrow 7 \rightarrow 6 \rightarrow 5 \rightarrow 4 \rightarrow 12 \rightarrow 14 \rightarrow 15 \rightarrow 13$

2. Determining path-specific penalty terms: the output is listed in Table 1.1.

Table 1.1. Path-specific $\Psi_{ip_{a(S\rightarrow R)}}$ values for the corresponding cells.

Path	Cell (bold) and $\Psi_{ip_{a(S\rightarrow R)}}$							
P1	**7**	**3**	**2**	**1**	R			
	2	3	4	5	6			
P2	**7**	**3**	**2**	**1**	**8**	**9**		
	2	3	4	5	6	7		
P3	**7**	**3**	**2**	**1**	**8**	**10**	**11**	**9**
	2	3	4	5	6	7	8	9
P4	**7**	**6**	**5**	**4**	R			
	2	3	4	5	6			
P5	**7**	**6**	**5**	**4**	**12**	**13**		
	2	3	4	5	6	7		
P6	**7**	**6**	**5**	**4**	**12**	**14**	**15**	**13**
	2	3	4	5	6	7	8	9

3. Selecting the cell-wise $\Psi_i = \max(\Psi_{ip_{a(S\rightarrow R)}})$ from Table 1.1: (output in Table 1.2)

Table 1.2. The finalized cell-specific Ψ_i values (also shown in Figure A.2)

Cell→	7	3	2	1	R	8	9	10	11	6	5	4	12	13	14	15
Ψ_i→	2	3	4	5	6	6	9	7	8	3	4	5	6	9	7	8

End

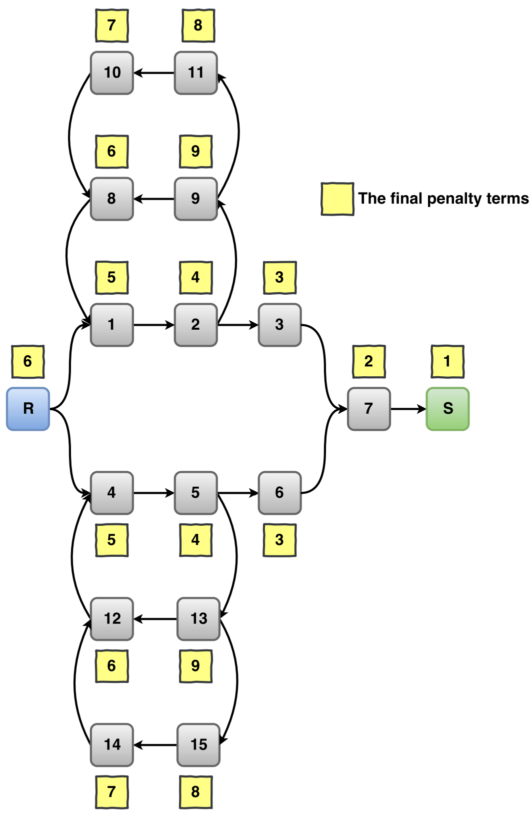


Figure A.2. The determined cell-specific Ψ_i values are shown in the yellow boxes.

Appendix B

B.1 Generating Ψ_{i_B}s and Ψ_{i_G}s: An Example

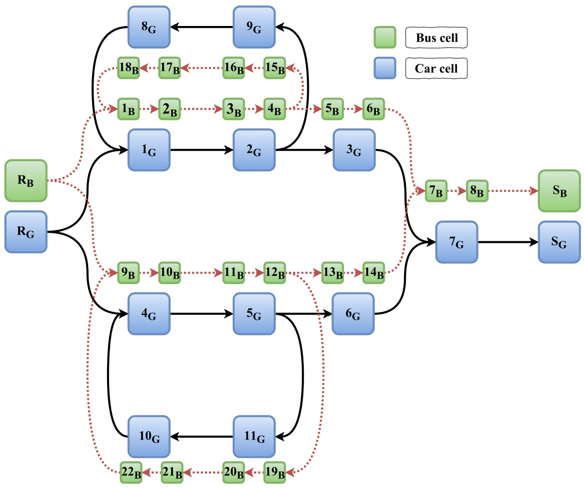


Figure B.1. The example network for which Ψ_{i_B} and Ψ_{i_G} values are determined. There are four cyclic loops in the network. Two for bus paths and two for car paths. The network is symmetric with respect to the corresponding type of source and destination cells. For this example we assume $L_B = 1$km and $L_G = 2$km.

The algorithm was presented in Section 5.2.1.1. For the network in Figure B.1, we outline the output of the algorithm below.

Start

1. Initialization:

   (a) $L_B = 1$km, $L_G = 2$km.

   (b) $\Psi_{S_B} = 0$, $\Psi_{S_G} = 0$.

(c) Determining set of all the bus paths $\mathcal{P}_B$:

i. Bus path-1 ($P_{1S_B \to R_B}$): $S_B \to 8_B \to 7_B \to 6_B \to 5_B \to 4_B \to 3_B \to 2_B \to 1_B \to R_B$

ii. Bus path-2 ($P_{2S_B \to R_B}$): $S_B \to 8_B \to 7_B \to 6_B \to 5_B \to 4_B \to 3_B \to 2_B \to 1_B \to 18_B \to 17_B \to 16_B \to 15_B$

iii. Bus path-3 ($P_{3S_B \to R_B}$): $S_B \to 8_B \to 7_B \to 14_B \to 13_B \to 12_B \to 11_B \to 10_B \to 9_B \to R_B$

iv. Bus path-4 ($P_{4S_B \to R_B}$): $S_B \to 8_B \to 7_B \to 14_B \to 13_B \to 12_B \to 11_B \to 10_B \to 9_B \to 22_B \to 21_B \to 20_B \to 19_B$

(d) Determining set of all the car paths $\mathcal{P}_G$:

i. Car path-1 ($P_{1S_G \to R_G}$): $S_G \to 7_G \to 3_G \to 2_G \to 1_G \to R_G$

ii. Car path-2 ($P_{2S_G \to R_G}$): $S_G \to 7_G \to 3_G \to 2_G \to 1_G \to 8_G \to 9_G$

iii. Car path-3 ($P_{3S_G \to R_G}$): $S_G \to 7_G \to 6_G \to 5_G \to 4_G \to R_G$

iv. Car path-4 ($P_{4S_G \to R_G}$): $S_G \to 7_G \to 6_G \to 5_G \to 4_G \to 10_G \to 11_G$

2. Determining path-specific penalty terms for the bus cells: the output is listed in Table 2.1.

Table 2.1. Path-specific penalty values for the corresponding cells.

Paths	Cells and penalty terms											
$P_{1S_B \to R_B}$	8_B	7_B	6_B	5_B	4_B	3_B	2_B	1_B	R_B			
	1	2	3	4	5	6	7	8	9			
$P_{2S_B \to R_B}$	8_B	7_B	6_B	5_B	4_B	3_B	2_B	1_B	18_B	17_B	16_B	15_B
	1	2	3	4	5	6	7	8	9	10	11	12
$P_{3S_B \to R_B}$	8_B	7_B	14_B	13_B	12_B	11_B	10_B	9_B	R_B			
	1	2	3	4	5	6	7	8	9			
$P_{4S_B \to R_B}$	8_B	7_B	14_B	13_B	12_B	11_B	10_B	9_B	22_B	21_B	20_B	19_B
	1	2	3	4	5	6	7	8	9	10	11	12
$P_{1S_G \to R_G}$	7_G	3_G	2_G	1_G	R_G							
	2	4	6	8	9							
$P_{2S_G \to R_G}$	7_G	3_G	2_G	1_G	8_G	9_G						
	2	4	6	8	10	12						
$P_{3S_G \to R_G}$	7_G	6_G	5_G	4_G	R_G							
	2	4	6	8	9							
$P_{4S_G \to R_G}$	7_G	6_G	5_G	4_G	10_G	11_G						
	2	4	6	8	10	12						

3. Selecting the cell-wise $\Psi_{i_B} = \max(\Psi_{i_B p_{a(S_B \to R_B)}})$ and $\Psi_{i_G} = \max(\Psi_{i_G p_{a(S_G \to R_G)}})$ from Table 2.1: (output in Table 2.2)

Table 2.2. Cell specific penalty terms for the car and bus cells.

Bus cells→	8_B	7_B	6_B	5_B	4_B	3_B	2_B	1_B	R_B	18_B	17_B	16_B	15_B
Ψ_{i_B} →	1	2	3	4	5	6	7	8	9	9	10	11	12
Bus cells→	14_B	13_B	12_B	11_B	10_B	9_B	22_B	21_B	20_B	19_B			
Ψ_{i_B} →	3	4	5	6	7	8	9	10	11	12			
Car Cells→	7_G	3_G	2_G	1_G	R_G	8_G	9_G	6_G	5_G	4_G	10_G	11_G	
Ψ_{i_G} →	2	4	6	8	9	10	12	4	6	8	10	12	

4. Normalize with respect to $\Psi_{max} = 12$. Results are presented in Table 2.3

Table 2.3. Normalized penalty terms.

Bus cells→	8_B	7_B	6_B	5_B	4_B	3_B	2_B	1_B	R_B	18_B	17_B	16_B
Ψ_{i_B} →	0.08	0.17	0.25	0.33	0.42	0.5	0.58	0.67	0.75	0.75	0.83	0.92
Bus cells→	15_B	14_B	13_B	12_B	11_B	10_B	9_B	22_B	21_B	20_B	19_B	
Ψ_{i_B} →	1	0.25	0.33	0.42	0.5	0.58	0.67	0.75	0.83	0.92	1	
Car cells→	7_G	3_G	2_G	1_G	R_G	8_G	9_G	6_G	5_G	4_G	10_G	11_G
Ψ_{i_G} →	0.17	0.33	0.5	0.67	0.75	0.83	1	0.33	0.5	0.67	0.83	1

End

B.2 NHB Proof of the BP SO-DTA-SC Model

Proposition B.2.1. *The optimal solution of the proposed BP SO-DTA-SC formulation does not have any HB flow.*

Proof. Suppose that the optimal solution $(\bar{x}, \bar{y})$ of the proposed BP SO-DTA-SC formulation holds back traffic somewhere within the network. Due to the inclusion of the equality constraint (5.49) in the formulation, both of these solutions attain SO TSPT. We also assume, a cell connector (i_B, j_B) belonging to an ordinary bus cell is involved in the HB process. Let the HB occurs at cell i_B and Δ denote the amount of HB traffic in cell i_B at time slot $t+1$. Then, one of the following strict inequalities for the $\mathcal{E}_{O_B}$ must hold at the time slot t.

$$\exists(i_B, j_B) \in \mathcal{E}_{O_B},\ \bar{y}^t_{i_B j_B} < \min\left\{\bar{x}^t_{i_B},\ Q^t_{i_B},\ Q^t_{j_B}, \delta^t_{j_B}\left(N^t_{j_B} - \bar{x}^t_{j_B}\right)\right\}. \tag{B.1}$$

Therefore, it is possible to move Δ amount of traffic from the ordinary bus cell i_B to some downstream bus cell j_B, at time t, and keep all other bus and car occupancies unchanged. Let $(x^\star,\ y^\star)$ be the NHB solution obtained by moving this HB traffic and $\bar{\mathcal{O}}$ and $\mathcal{O}^\star$ denote the objective values for the solutions $(\bar{x}, \bar{y})$ and $(x^\star, y^\star)$, respectively. Then, one can write

$$\begin{aligned}
\mathcal{O}^\star - \bar{\mathcal{O}} &= \sum_{\forall t\in\mathcal{T}} \sum_{\forall i\in\mathcal{C}_B\setminus\mathcal{C}_{S_B}} \alpha_B \cdot \Psi_{i_B} \cdot x^{\star t}_{i_B} - \sum_{\forall t\in\mathcal{T}} \sum_{\forall i\in\mathcal{C}_B\setminus\mathcal{C}_{S_B}} \alpha_B \cdot \Psi_{i_B} \cdot \bar{x}^t_{i_B} \\
&= \alpha_B\left(\Psi_{i_B} x^{\star t}_{i_B} + \Psi_{j_B} x^{\star t+1}_{j_B} - \Psi_{i_B}\bar{x}^t_{i_B} - \Psi_{j_B}\bar{x}^{t+1}_{j_B}\right) \\
&= \alpha_B\{\Psi_{j_B}\left(x^{\star t+1}_{j_B} - \bar{x}^{t+1}_{j_B}\right) - \Psi_{i_B}\left(\bar{x}^t_{i_B} - x^{\star t}_{i_B}\right)\} \\
&= \alpha_B\left(\Psi_{j_B}\Delta - \Psi_{i_B}\Delta\right) \\
&= \alpha_B\left(\Psi_{j_B} - \Psi_{i_B}\right)\Delta.
\end{aligned} \tag{B.2}$$

Since $\alpha_B,\ \Delta > 0$, we have $\mathcal{O}^\star < \bar{\mathcal{O}}$, because $\Psi_{i_B} > \Psi_{j_B}$. This cannot happen as the $\bar{\mathcal{O}}$ is, by definition, the optimal objective value of the BP SO-DTA-SC minimization objective in Equation (5.1). As a result, Δ must be zero, which implies that there is no holding back in the solution of the proposed BP SO-DTA-SC formulation. The proof for the general case is a straightforward extension and involves moving the Δ amount of traffic between the multiple cells of the corresponding bus or car traffic class. $\square$

B.3 A Mixed-integer Transit Signal Priority (MTSP) Formulation

Let us consider mixed-integer signal control variables for bus and car Υ^t_B and Υ^t_G, respectively. Then one can write following flow control constraint,

$$y^t_{(z,p)_\ell j_\ell} \leqslant \Upsilon^t_{(z,p)_\ell} Q_{(z,p)_\ell}, \ell \in \{B, G\},\ z \in \mathcal{I},\ p \in \{1,2,3,4\},\ (z,p) \in \mathcal{C}_I,\ \forall t \in \mathcal{T}. \tag{B.3}$$

Similar to the approach presented by Liu et al. (2015a), bus-indicator and bus-priority of TSP at an intersection can be modelled using the following equations:

$$x^t_{(z,p)_B} \geqslant \Upsilon^t_{(z,p)_B},\ z \in \mathcal{I},\ p \in \{1,2,3,4\},\ (z,p) \in \mathcal{C}_I,\ \forall t \in \mathcal{T}, \tag{B.4}$$

$$x^t_{(z,p)_B} \leqslant \Lambda \cdot \Upsilon^t_{(z,p)_B},\ z \in \mathcal{I},\ p \in \{1,2,3,4\},\ (z,p) \in \mathcal{C}_I,\ \forall t \in \mathcal{T}, \tag{B.5}$$

$$\Upsilon^t_{(z,p)_B} + \Upsilon^t_{(z,p)_G} = 1,\ z \in \mathcal{I},\ p \in \{1,2,3,4\},\ (z,p) \in \mathcal{C}_I,\ \forall t \in \mathcal{T}, \tag{B.6}$$

where Λ is a very large number and parameter of the model. The above three equations (B.4)-(B.6) assure that when a bus arrives at the intersection, it gets priority over waiting cars. One can model signal coordination and conflict avoidance constrains for MTSP by following quations (5.39)-(5.48) and are omitted here.

Bibliography

Ahn, Kyoungho. 2002. “Modeling Light Duty Vehicle Emissions Based on Instantaneous Speed and Acceleration Levels.” PhD diss.

Al-Deek, Haitham, Adrian Sandt, Ahmad Alomari, and Omar Hussain. 2017. “A technical note on evaluating the effectiveness of bus rapid transit with transit signal priority.” *Journal of Intelligent Transportation Systems* 21 (3): 227–238.

Alonso, Borja, Juan Carlos Munoz, Angel Ibeas, and Jose Luis Moura. 2017. “A congested and dwell time dependent transit corridor assignment model.” *Journal of Advanced Transportation* .

Aziz, HM, and Satish V Ukkusuri. 2012a. “Integration of environmental objectives in a system optimal dynamic traffic assignment model.” *Computer-Aided Civil and Infrastructure Engineering* 27 (7): 494–511.

Aziz, HM, and Satish V Ukkusuri. 2012b. “Integration of environmental objectives in a system optimal dynamic traffic assignment model.” *Computer-Aided Civil and Infrastructure Engineering* 27 (7): 494–511.

Aziz, H.M.A., and S.V. Ukkusuri. 2011. “An analytical framework for vehicular traffic signal control integrated with dynamic traffic assignment using cell transmission model.” In *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*, Oct, 1355–1360.

Aziz, H.M.A., and S.V. Ukkusuri. 2012c. “Unified Framework for Dynamic Traffic Assignment and Signal Control with Cell Transmission Model.” *Transportation Research Record: Journal of the Transportation Research Board* 2311 (1): 73–84.

Bagloee, Saeed Asadi, and Majid Sarvi. 2015. “Heuristic approach to capacitated traffic assignment problem for large-scale transport networks.” *Transportation Research Record: Journal of the Transportation Research Board* (2498): 1–11.

Balijepalli, NC, D Ngoduy, and DP Watling. 2014. “The two-regime transmission model for network loading in dynamic traffic assignment problems.” *Transportmetrica A: Transport Science* 10 (7): 563–584.

Beard, C., and A. Ziliaskopoulos. 2006. "System Optimal Signal Optimization Formulation." *Transportation Research Record: Journal of the Transportation Research Board* 1978 (1): 102–112.

Bell, Michael GH, Caroline M Shield, Fritz Busch, and Gunter Kruse. 1997. "A stochastic user equilibrium path flow estimator." *Transportation Research Part C: Emerging Technologies* 5 (3): 197–210.

Benedek, Christine M, and Laurence R Rilett. 1998. "Equitable traffic assignment with environmental cost functions." *Journal of transportation engineering* 124 (1): 16–22.

Bliemer, Michiel. 2007. "Dynamic queuing and spillback in analytical multiclass dynamic network loading model." *Transportation Research Record: Journal of the Transportation Research Board* (2029): 14–21.

Boyce, David, and Qian Xiong. 2004. "User-optimal and system-optimal route choices for a large road network." *Review of network Economics* 3 (4).

Boyd, Stephen, and Lieven Vandenberghe. 2004. *Convex optimization.* Cambridge university press.

Carey, Malachy, and Eswaran Subrahmanian. 2000. "An approach to modelling time-varying flows on congested networks." *Transportation Research Part B: Methodological* 34 (3): 157–183.

Carey, Malachy, and David Watling. 2012. "Dynamic traffic assignment approximating the kinematic wave model: System optimum, marginal costs, externalities and tolls." *Transportation Research Part B: Methodological* 46 (5): 634–648.

Chang, Gang-len, Meenakshy Vasudevan, and Chih-chiang Su. 1996. "Modelling and evaluation of adaptive bus-preemption control with and without automatic vehicle location systems." *Transportation Research Part A: Policy and Practice* 30 (4): 251–268.

Chang, Liu, and Wang Hui. 2016. "Traffic emission control based on emission pricing and signal timing." In *Intelligent Control and Automation (WCICA), 2016 12th World Congress on*, 467–472. IEEE.

Chen, Owen Jianwen, and Moshe E. Ben-Akiva. 1998. "Game-Theoretic Formulations of Interaction Between Dynamic Traffic Control and Dynamic Traffic Assignment." *Transportation Research Record: Journal of the Transportation Research Board* 1617 (1): 179 – 188.

Cheung, Landy L, and Amer S Shalaby. 2017. "System Optimal Re-Routing Transit Assignment Heuristic: A Theoretical Framework and Large-Scale Case Study." *International Journal of Transportation Science and Technology* .

Chiang, Mung, Steven H Low, A Robert Calderbank, and John C Doyle. 2007. “Layering as optimization decomposition: A mathematical theory of network architectures.” *Proceedings of the IEEE* 95 (1): 255–312.

Chiu, Yi-Chang, Jon Bottom, Michael Mahut, Alex Paz, Ramachandran Balakrishna, Travis Waller, and Jim Hicks. 2011. “Dynamic traffic assignment: A primer.” *Transportation Research E-Circular* (E-C153).

Chiu, Yi-Chang, and Hong Zheng. 2007. “Real-time mobilization decisions for multi-priority emergency response resources and evacuation groups: model formulation and solution.” *Transportation Research Part E: Logistics and Transportation Review* 43 (6): 710–736.

Chow, Andy HF. 2009. “Properties of system optimal traffic assignment with departure time choice and its solution method.” *Transportation Research Part B: Methodological* 43 (3): 325–344.

Como, Giacomo, Enrico Lovisari, and Ketan Savla. 2016. “Convexity and robustness of dynamic network traffic assignment for control of freeway networks.” *IFAC-PapersOnLine* 49 (3): 335–340.

Daganzo, Carlos 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, Carlos F. 1995. “The cell transmission model, part II: Network traffic.” *Transportation Research Part B: Methodological* 29 (2): 79 – 93.

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

Dirks, Kim N., Murray D. Johns, John E. Hay, and Andrew P. Sturman. 2003. “A semi-empirical model for predicting the effect of changes in traffic flow patterns on carbon monoxide concentrations.” *Atmospheric Environment* 37 (19): 2719 – 2724.

Doan, Kien, and Satish V. Ukkusuri. 2012. “On the holding-back problem in the cell transmission based dynamic traffic assignment models.” *Transportation Research Part B: Methodological* 46 (9): 1218 – 1238.

Doan, Kien, and Satish V Ukkusuri. 2015. “Dynamic system optimal model for multi-OD traffic networks with an advanced spatial queuing model.” *Transportation Research Part C: Emerging Technologies* 51: 41–65.

Dong, Xisong, Yuetong Lin, Dayong Shen, Zhengxi Li, Fenghua Zhu, Bin Hu, Dong Fan, and Gang Xiong. 2017. “A Parallel Transportation Management and Control System for Bus Rapid Transit Using the ACP Approach.” *IEEE Transactions on Intelligent Transportation Systems* .

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

Ernst, John. 2005. "Initiating bus rapid transit in Jakarta, Indonesia." *Transportation Research Record: Journal of the Transportation Research Board* (1903): 20–26.

Fernandes, P, MC Coelho, and NM Rouphail. 2017. "Assessing the impact of closely-spaced intersections on traffic operations and pollutant emissions on a corridor level." *Transportation Research Part D: Transport and Environment* 54: 304–320.

Friesz, Terry L, Ke Han, Hongcheng Liu, and Tao Yao. 2013. "Dynamic congestion and tolls with mobile source emission." *Procedia-Social and Behavioral Sciences* 80: 818–836.

Gan, Albert, Huanyu Yue, Ike Ubaka, and Fang Zhao. 2003. "Development of operational performance and decision models for arterial bus lanes." *Transportation Research Record: Journal of the Transportation Research Board* (1858): 18–30.

Gao, Kaizhou, Yicheng Zhang, Ali Sadollah, and Rong Su. 2017. "Improved artificial bee colony algorithm for solving urban traffic light scheduling problem." In *Evolutionary Computation (CEC), 2017 IEEE Congress on*, 395–402. IEEE.

Greenshields, BD, Ws Channing, Hh Miller, et al. 1935. "A study of traffic capacity." In *Highway research board proceedings*, Vol. 1935. National Research Council (USA), Highway Research Board.

Hajbabaie, Ali, and Rahim F Benekohal. 2015. "A program for simultaneous network signal timing optimization and traffic assignment." *IEEE Transactions on Intelligent Transportation Systems* 16 (5): 2573–2586.

He, Qing, K Head, and Jun Ding. 2011. "Heuristic algorithm for priority traffic signal control." *Transportation Research Record: Journal of the Transportation Research Board* (2259): 1–7.

He, Qing, K. Larry Head, and Jun Ding. 2014. "Multi-modal traffic signal control with priority, signal actuation and coordination." *Transportation Research Part C: Emerging Technologies* 46: 65 – 82.

Head, Larry, Douglas Gettman, and Zhiping Wei. 2006. "Decision model for priority control of traffic signals." *Transportation Research Record: Journal of the Transportation Research Board* (1978): 169–177.

Hoang, Nam H, Hai L Vu, Mike Smith, and Dong Ngoduy. 2016. "Convex signal control model in a single-destination dynamic traffic assignment." In *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*, 790–794. IEEE.

Hu, Xiaojian, Wei Wang, and Jian Lu. 2010. *Urban Short-Term Traffic Flow Forecasting Based on the Semi-Variable Cell Transmission Model*, Chap. 80, 861–871.

Islam, Tarikul, Hai L Vu, Manoj Panda, and Dong Ngoduy. 2017. "A study of realistic dynamic traffic assignment with signal control, time-scale, and emission." *Journal of Intelligent Transportation Systems* 1–16.

Janos, Melanie, and Peter Furth. 2002. "Bus priority with highly interruptible traffic signal control: simulation of San Juan's Avenida Ponce de Leon." *Transportation Research Record: Journal of the Transportation Research Board* (1811): 157–165.

Janson, Bruce N. 1991. "Dynamic traffic assignment for urban road networks." *Transportation Research Part B: Methodological* 25 (2): 143–161.

Jones, Lee K., Rahul Deshpande, Nathan H. Gartner, Chronis Stamatiadis, and Fei Zou. 2013. "Robust controls for traffic networks: The near-Bayes near-Minimax strategy." *Transportation Research Part C: Emerging Technologies* 27: 205 – 218. Selected papers from the Seventh Triennial Symposium on Transportation Analysis (TRISTAN VII).

Larsson, T, and M Patriksson. 1996. "Side constrained traffic equilibrium models: analysis, computation and applications, revised for possible publication in Transportation Research." *Part B* .

Larsson, Torbjörn, and Michael Patriksson. 1995. "An augmented lagrangean dual algorithm for link capacity side constrained traffic assignment problems." *Transportation Research Part B: Methodological* 29 (6): 433 – 455.

Le, Tung, Péter Kovács, Neil Walton, Hai L. Vu, Lachlan L.H. Andrew, and Serge S.P. Hoogendoorn. 2015. "Decentralized signal control for urban road networks." *Transportation Research Part C: Emerging Technologies* 58, Part C: 431 – 450. Special Issue: Advanced Road Traffic Control.

Lertworawanich, Ponlathep, Masao Kuwahara, and Marc Miska. 2011. "A new multiobjective signal optimization for oversaturated networks." *IEEE Transactions on Intelligent Transportation Systems* 12 (4): 967–976.

LeVeque, Randall J. 1992. *Numerical methods for conservation laws.* Vol. 132. Springer.

Levin, Michael, Melissa Duell, and S Waller. 2014. "Effect of Road Grade on Networkwide Vehicle Energy Consumption and Ecorouting." *Transportation Research Record: Journal of the Transportation Research Board* (2427): 26–33.

Li, ShuGuang, and YongFeng Ju. 2009. "Evaluation of bus-exclusive lanes." *IEEE Transactions on Intelligent Transportation Systems* 10 (2): 236–245.

Li, Yue, S Travis Waller, and Thanasis Ziliaskopoulos. 2003. "A decomposition scheme for system optimal dynamic traffic assignment models." *Networks and Spatial Economics* 3 (4): 441–455.

Li, Yue, Athanasios Ziliaskopoulos, and S Waller. 1999. "Linear programming formulations for system optimum dynamic traffic assignment with arrival time-based and departure time-based demands." *Transportation Research Record: Journal of the Transportation Research Board* (1667): 52–59.

Li, Yue, Thanasis Ziliaskopoulos, and David Boyce. 2002. "Combined model for time-dependent trip distribution and traffic assignment." *Transportation Research Record: Journal of the Transportation Research Board* (1783): 98–109.

Li, Z. 2011. "Modeling Arterial Signal Optimization with Enhanced Cell Transmission Formulations." *Journal of Transportation Engineering* 137 (7): 445–454.

Li, Zhi-Chun, Hai-Jun Huang, William HK Lam, and Sze Chun Wong. 2007. "A model for evaluation of transport policies in multimodal networks with road and parking capacity constraints." *Journal of Mathematical Modelling and Algorithms* 6 (2): 239–257.

Liao, Tsai-Yun. 2013. "A fuel-based signal optimization model." *Transportation Research Part D: Transport and Environment* 23 (0): 1 – 8.

Lin, Wei-Hua, and Chenghong Wang. 2004. "An enhanced 0-1 mixed-integer LP formulation for traffic signal control." *Intelligent Transportation Systems, IEEE Transactions on* 5 (4): 238–245.

Liu, Haiyang, Jian Wang, Kasun Wijayaratna, Vinayak V Dixit, and Steven Travis Waller. 2015a. "Integrating the Bus Vehicle Class Into the Cell Transmission Model." *IEEE Transactions on Intelligent Transportation Systems* 16 (5): 2620–2630.

Liu, Huiyuan, Lihui Zhang, Daniel Sun, and Dianhai Wang. 2015b. "Optimize the settings of variable speed limit system to improve the performance of freeway traffic." *IEEE Transactions on Intelligent Transportation Systems* 16 (6): 3249–3257.

Liu, Yue, and Gang-Len Chang. 2011. "An arterial signal optimization model for intersections experiencing queue spillback and lane blockage." *Transportation research part C: emerging technologies* 19 (1): 130–144.

Lo, Hong K. 1999. "A novel traffic signal control formulation." *Transportation Research Part A: Policy and Practice* 33 (6): 433 – 448.

Lo, Hong K. 2001. "A Cell-Based Traffic Control Formulation: Strategies and Benefits of Dynamic Timing Plans." *Transportation Science* 35 (2): 148–164.

Lo, Hong K., Elbert Chang, and Yiu Cho Chan. 2001. “Dynamic network traffic control.” *Transportation Research Part A: Policy and Practice* 35 (8): 721 – 744.

Long, Jiancheng, Jiaxu Chen, WY Szeto, and Qin Shi. 2016. “Link-based system optimum dynamic traffic assignment problems with environmental objectives.” *Transportation Research Part D: Transport and Environment* .

Lu, Chung-Cheng, Jiangtao Liu, Yunchao Qu, Srinivas Peeta, Nagui M Rouphail, and Xuesong Zhou. 2016. “Eco-system optimal time-dependent flow assignment in a congested network.” *Transportation Research Part B: Methodological* 94: 217–239.

Luo, Lihua, Ying-En Ge, Fangwei Zhang, and Xuegang Jeff Ban. 2016. “Real-time route diversion control in a model predictive control framework with multiple objectives: Traffic efficiency, emission reduction and fuel economy.” *Transportation Research Part D: Transport and Environment* 48: 332–356.

Ma, Rui, Xuegang Jeff Ban, and WY Szeto. 2015. “Emission modeling and pricing in dynamic traffic networks.” *Transportation Research Procedia* 9: 106–129.

Ma, Rui, Xuegang Jeff Ban, and WY Szeto. 2017. “Emission modeling and pricing on single-destination dynamic traffic networks.” *Transportation Research Part B: Methodological* 100: 255–283.

Ma, Wanjing, Yue Liu, and Xiaoguang Yang. 2013. “A dynamic programming approach for optimal signal priority control upon multiple high-frequency bus requests.” *Journal of Intelligent Transportation Systems* 17 (4): 282–293.

Ma, Wanjing, Xiaoguang Yang, and Yue Liu. 2010. “Development and evaluation of a coordinated and conditional bus priority approach.” *Transportation Research Record: Journal of the Transportation Research Board* (2145): 49–58.

Mahmassani, Hani S. 2001. “Dynamic network traffic assignment and simulation methodology for advanced system management applications.” *Networks and spatial economics* 1 (3): 267–292.

Mehrabipour, Mehrzad, and Ali Hajbabaie. 2017. “A Cell-Based Distributed-Coordinated Approach for Network-Level Signal Timing Optimization.” *Computer-Aided Civil and Infrastructure Engineering* 32 (7): 599–616.

Mensink, C, I De Vlieger, and J Nys. 2000. “An urban transport emission model for the Antwerp area.” *Atmospheric Environment* 34 (27): 4595 – 4602.

Merchant, Deepak K, and George L Nemhauser. 1978a. “A model and an algorithm for the dynamic traffic assignment problems.” *Transportation science* 12 (3): 183–199.

Merchant, Deepak K, and George L Nemhauser. 1978b. “Optimality conditions for a dynamic traffic assignment model.” *Transportation Science* 12 (3): 200–207.

Mesa-Arango, Rodrigo, and Satish V Ukkusuri. 2014. "Modeling the car-truck interaction in a system-optimal dynamic traffic assignment model." *Journal of Intelligent Transportation Systems* 18 (4): 327–338.

Messmer, A, and M Papageorgiou. 2000. "METANET-A simulation program for motorway networks." *Technical University of Crete, Dynamic Systems and Simulation Laboratory* .

Munoz, L., Xiaotian Sun, R. Horowitz, and L. Alvarez. 2003. "Traffic density estimation with the cell transmission model." In *American Control Conference, 2003. Proceedings of the 2003*, Vol. 5, June, 3750–3755 vol.5.

Nie, Xiaojian, and H Michael Zhang. 2005. "A comparative study of some macroscopic link models used in dynamic traffic assignment." *Networks and Spatial Economics* 5 (1): 89–115.

Nie, Yu. 2006. "A variational inequality approach for inferring dynamic origin-destination travel demands." PhD diss., University of California Davis.

Nie, Yu, HM Zhang, and Der-Horng Lee. 2004. "Models and algorithms for the traffic assignment problem with link capacity constraints." *Transportation Research Part B: Methodological* 38 (4): 285–312.

Palomar, Daniel Pérez, and Mung Chiang. 2006. "A tutorial on decomposition methods for network utility maximization." *IEEE Journal on Selected Areas in Communications* 24 (8): 1439–1451.

Pasquale, C, S Sacone, S Siri, and B De Schutter. 2017. "A multi-class model-based control scheme for reducing congestion and emissions in freeway networks by combining ramp metering and route guidance." *Transportation Research Part C: Emerging Technologies* 80: 384–408.

Peeta, Srinivas, and Athanasios K Ziliaskopoulos. 2001. "Foundations of dynamic traffic assignment: The past, the present and the future." *Networks and Spatial Economics* 1 (3): 233–265.

Pipes, Louis A. 1967. "Car following models and the fundamental diagram of road traffic." *Transportation Research* 1 (1): 21–29.

Pisarski, Dominik, and Carlos Canudas-de Wit. 2012. "Optimal balancing of road traffic density distributions for the cell transmission model." In *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, 6969–6974. IEEE.

Pohlmann, T., and B. Friedrich. 2010. "Online control of signalized networks using the Cell Transmission Model." In *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*, Sept, 1–6.

Qian, Zhen Sean, Wei Shen, and HM Zhang. 2012. "System-optimal dynamic traffic assignment with and without queue spillback: Its path-based formulation and solution via approximate path marginal cost." *Transportation research part B: methodological* 46 (7): 874–893.

Ren, Hualing, Haoxiang Liu, Jiancheng Long, and Ziyou Gao. 2013. "Dynamic user optimal signal design at isolated intersections." *PROMET-Traffic&Transportation* 25 (1): 13–22.

Robert, C. 2013. "Bus Rapid Transit (BRT): An Efficient and Competitive Mode of Public Transport." *20th ACEA Scientific and Advisory Group (SAG) report* 1–36.

Ruan, Jin-Mei, Bin Liu, He Wei, Yunchao Qu, Nana Zhu, and Xuesong Zhou. 2016. "How many and where to locate parking lots? A space–time accessibility-maximization modeling framework for special event traffic management." *Urban Rail Transit* 2 (2): 59–70.

Satiennam, Thaned, Atsushi Fukuda, and Ryosuke Oshima. 2006. "A study on the introduction of bus rapid transit system in Asian developing cities: A case study on Bangkok Metropolitan Administration Project." *IATSS research* 30 (2): 59–69.

Sayegh, Arwa. 2017. "Uncertainties and Errors in Predicting Vehicle Exhaust Emissions using Traffic Flow Models." PhD diss., University of Leeds.

Sayegh, Arwa S, Richard D Connors, and James E Tate. 2017. "Uncertainty propagation from the cell transmission traffic flow model to emission predictions: a data-driven approach." *Transportation Science* .

Shahpar, Amir H., Hedayat Z. Aashtiani, and Abbas Babazadeh. 2008. "Dynamic penalty function method for the side constrained traffic assignment problem." *Applied Mathematics and Computation* 206 (1): 332 – 345.

Shakkottai, Srinivas, Rayadurgam Srikant, et al. 2008. "Network optimization and control." *Foundations and Trends® in Networking* 2 (3): 271–379.

Shen, Wei, Yu Nie, and H Zhang. 2007. "Dynamic network simplex method for designing emergency evacuation plans." *Transportation Research Record: Journal of the Transportation Research Board* (2022): 83–93.

Shen, Wei, and HM Zhang. 2014. "System optimal dynamic traffic assignment: Properties and solution procedures in the case of a many-to-one network." *Transportation Research Part B: Methodological* 65: 1–17.

Smaili, Samia, Salim Mammar, and Said Mammar. 2011. "Cell transmission model and numerical schemes for mixed traffic." *Logistics (LOGISTIQUA), 2011 4th International Conference on* 268–273.

Smith, MJ. 1980. “A local traffic control policy which automatically maximises the overall travel capacity of an urban road network.” *Traffic Engineering & Control* 21 (HS-030 129).

Sugawara, Satoshi, and D Niemeier. 2002. “How much can vehicle emissions be reduced?: exploratory analysis of an upper boundary using an emissions-optimized trip assignment.” *Transportation Research Record: Journal of the Transportation Research Board* (1815): 29–37.

Sumalee, A, RX Zhong, TL Pan, and WY Szeto. 2011. “Stochastic cell transmission model (SCTM): A stochastic dynamic traffic model for traffic state surveillance and assignment.” *Transportation Research Part B: Methodological* 45 (3): 507–533.

Sun, D., and A. Bayen. 2008. “Multicommodity Eulerian-Lagrangian Large-capacity Cell Transmission Model for en route traffic.” In *AIAA Journal of Guidance, Control and Dynamics*, Vol. 31, June, 616–6128 No. 3.

Teklu, Fitsum, Agachai Sumalee, and David Watling. 2007. “A genetic algorithm approach for optimizing traffic control signals considering routing.” *Computer-Aided Civil and Infrastructure Engineering* 22 (1): 31–43.

The MathWorks, Inc. 2017. “Confidence and Prediction Bounds.” jun. https://au.mathworks.com/help/curvefit/confidence-and-prediction-bounds.html.

Timotheou, Stelios, Christos G Panayiotou, and Marios M Polycarpou. 2014. “Online distributed network traffic signal control using the cell transmission model.” In *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*, 2523–2528. IEEE.

Timotheou, Stelios, Christos G Panayiotou, and Marios M Polycarpou. 2015. “Distributed traffic signal control using the cell transmission model via the alternating direction method of multipliers.” *IEEE Transactions on Intelligent Transportation Systems* 16 (2): 919–933.

Tokuda, Sho, Ryo Kanamori, and Takayuki Ito. 2017. “A Modification of the Stochastic Cell Transmission Model for Urban Networks.” *International Journal of Intelligent Transportation Systems Research* 15 (2): 73–84.

Ukkusuri, Satish V., Gitakrishnan Ramadurai, and Gopal Patil. 2010. “A robust transportation signal control problem accounting for traffic dynamics.” *Computers & Operations Research* 37 (5): 869 – 879. Disruption Management.

Varaiya, Pravin. 2013. “The max-pressure controller for arbitrary networks of signalized intersections.” In *Advances in Dynamic Network Modeling in Complex Transportation Systems*, 27–66. Springer.

Varia, HR, and SL Dhingra. 2004. "Dynamic optimal traffic assignment and signal time optimization using genetic algorithms." *Computer-Aided Civil and Infrastructure Engineering* 19 (4): 260–273.

Viegas, José, and Baichuan Lu. 2004. "The intermittent bus lane signals setting within an area." *Transportation Research Part C: Emerging Technologies* 12 (6): 453–469.

Waller, S, and Athanasios Ziliaskopoulos. 2001. "Stochastic dynamic network design problem." *Transportation Research Record: Journal of the Transportation Research Board* (1771): 106–113.

Waller, S Travis, Kyriacos C Mouskos, Dimitrios Kamaryiannis, and Athanasios K Ziliaskopoulos. 2006. "A linear model for the continuous network design problem." *Computer-Aided Civil and Infrastructure Engineering* 21 (5): 334–345.

Waller, S. Travis, and Athanasios K. Ziliaskopoulos. 2006. "A chance-constrained based stochastic dynamic traffic assignment model: Analysis, formulation and solution algorithms." *Transportation Research Part C: Emerging Technologies* 14 (6): 418 – 427.

Wang, Jing, Xiaoning Zhang, and HM Zhang. 2016. "Parking permits management and optimal parking supply considering traffic emission cost." *Transportation Research Part D: Transport and Environment* .

Wang, Ping, L.S. Jones, Qun Yang, and S. Gurupackiam. 2013. "Cell transmission model based traffic signal timing in oversaturated conditions." *Journal of Central South University* 20 (4): 1129–1136.

Webster, FV. 1958. "Traffic signal settings, road research technical paper no. 39." *Road Research Laboratory* .

Wie, Byung-Wook, Roger L Tobin, and Terry L Friesz. 1994. "The augmented Lagrangian method for solving dynamic network traffic assignment models in discrete time." *Transportation Science* 28 (3): 204–220.

Wu, Jianping, and Nick Hounsell. 1998. "Bus priority using pre-signals." *Transportation Research Part A: Policy and Practice* 32 (8): 563–583.

Xia, Liping, and Yaping Shao. 2005. "Modelling of traffic flow and air pollution emission with application to Hong Kong Island." *Environmental Modelling & Software* 20 (9): 1175–1188.

Xing, Tao, and Xuesong Zhou. 2011. "Finding the most reliable path with and without link travel time correlation: A Lagrangian substitution based approach." *Transportation Research Part B: Methodological* 45 (10): 1660–1679.

Xing, Tao, and Xuesong Zhou. 2013. "Reformulation and solution algorithms for absolute and percentile robust shortest path problems." *IEEE Transactions on Intelligent Transportation Systems* 14 (2): 943–954.

Yperman, Isaak, Steven Logghe, Chris MJ Tampere, and Ben Immers. 2006. *Multicommodity link transmission model for dynamic network loading.* Technical report.

Zegeye, Solomon K, Bart De Schutter, Hans Hellendoorn, and Ewald Breunesse. 2009. "Model-based traffic control for balanced reduction of fuel consumption, emissions, and travel time." *IFAC Proceedings Volumes* 42 (15): 149–154.

Zhang, L., Y. Yin, and Y. Lou. 2010. "Robust Signal Timing for Arterials Under Day-to-Day Demand Variations." *Transportation Research Record: Journal of the Transportation Research Board* 2192 (1): 156 – 166.

Zhang, Yunlong, Jinpeng Lv, and Qi Ying. 2010. "Traffic assignment considering air quality." *Transportation Research Part D: Transport and Environment* 15 (8): 497–502.

Zhao, Junfeng, Wan Li, Junmin Wang, and Xuegang Ban. 2016. "Dynamic traffic signal timing optimization strategy incorporating various vehicle fuel consumption characteristics." *IEEE Transactions on Vehicular Technology* 65 (6): 3874–3887.

Zheng, Hong, and Yi-Chang Chiu. 2011. "A network flow algorithm for the cell-based single-destination system optimal dynamic traffic assignment problem." *Transportation Science* 45 (1): 121–137.

Zhong, RX, FF Yuan, TL Pan, Andy HF Chow, CJ Chen, and Z Yang. 2016. "Linear complementarity system approach to macroscopic freeway traffic modelling: uniqueness and convexity." *Transportmetrica A: Transport Science* 12 (2): 142–174.

Zhou, Li, Yizhe Wang, and Yangdong Liu. 2017. "Active Signal Priority Control Method for Bus Rapid Transit Based on Vehicle Infrastructure Integration." *International Journal of Transportation Science and Technology* .

Zhou, Xuesong, Shams Tanvir, Hao Lei, Jeffrey Taylor, Bin Liu, Nagui M. Rouphail, and H. Christopher Frey. 2015. "Integrating a simplified emission estimation model and mesoscopic dynamic traffic simulator to efficiently evaluate emission impacts of traffic management strategies." *Transportation Research Part D: Transport and Environment* 37: 123 – 136. http://www.sciencedirect.com/science/article/pii/S1361920915000450.

Zhu, Feng, Hong K. Lo, and Hong-Zhi Lin. 2013. "Delay and emissions modelling for signalised intersections." *Transportmetrica B: Transport Dynamics* 1 (2): 111–135.

Zhu, Feng, and Satish V. Ukkusuri. 2013. "A cell based dynamic system optimum model with non-holding back flows." *Transportation Research Part C: Emerging Technologies* 36: 367 – 380.

Zhu, Feng, and Satish V Ukkusuri. 2014. "Accounting for dynamic speed limit control in a stochastic traffic environment: A reinforcement learning approach." *Transportation research part C: emerging technologies* 41: 30–47.

Ziliaskopoulos, Athanasios K. 2000. "A Linear Programming Model for the Single Destination System Optimum Dynamic Traffic Assignment Problem." *Transportation Science* 34 (1): 37–49.