

Dynamic Traffic Congestion Pricing Mechanism with User-Centric Considerations

by

Kim Thien Bui

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

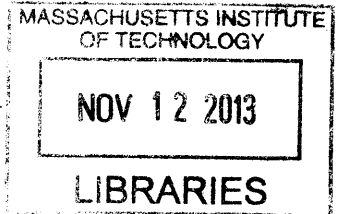
Master of Science in Transportation

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2013

ARCHIVES



© Massachusetts Institute of Technology 2013. All rights reserved.

Author
Department of Civil and Environmental Engineering
August 9, 2013

Certified by
Emilio Frazzoli
Professor of Aeronautics and Astronautics
Thesis Supervisor

Accepted by
Heidi Nepf
Chair, Committee for Graduate Students

Dynamic Traffic Congestion Pricing Mechanism with User-Centric Considerations

by

Kim Thien Bui

Submitted to the Department of Civil and Environmental Engineering
on August 9, 2013, in partial fulfillment of the
requirements for the degree of
Master of Science in Transportation

Abstract

In this thesis, we consider the problem of designing real-time traffic routing systems in urban areas. Optimal dynamic routing for multiple passengers is known to be computationally hard due to its combinatorial nature. To overcome this difficulty, we propose a novel mechanism called User-Centric Dynamic Pricing (UCDP) based on recent advances in algorithmic mechanism design. The mechanism allows for congestion-free traffic in general road networks with heterogeneous users, while satisfying each user's travel preference. The mechanism first informs whether a passenger should use public transportation or the road network. In the latter case, a passenger reports his maximum accepted travel time with a lower bound announced publicly by the road authority. The mechanism then assigns the passenger a path that matches with his preference given the current traffic condition in the network. The proposed mechanism introduces a fairness constrained shortest path (FCSP) problem with a special structure, thus enabling polynomial time computation of path allocation that maximizes the sequential social surplus and guarantees fairness among passengers. The tolls of paths are then computed according to marginal cost payments. We show that reporting true preference is a weakly dominant strategy. The performance of the proposed mechanism is demonstrated on several simulated routing experiments in comparison to user equilibrium and system optimum.

Thesis Supervisor: Emilio Frazzoli
Title: Professor of Aeronautics and Astronautics

Acknowledgments

This thesis would not have been possible without the guidance, support and help of many individuals to whom I want to express my sincere gratitude here.

First and foremost, my greatest appreciation goes to my thesis supervisor, Prof. Emilio Frazzoli, for his vision and guidance. He is truly a knowledgeable and insightful supervisor, who kept me focused on the big picture of problems and challenged me with his intelligent questions. At the same time, he encouraged me to work smartly and provided me with the freedom to explore fundamental research problems. He also gave me precious advices on how to manage work-life balance from his personal experience, which made my life at MIT so much easier.

I would like to thank Singapore-MIT Alliance for Research and Technology Center (SMART) for funding my research and hosting me during summers 2011 and 2012, when I conducted part of my research there. My gratitude also goes to outstanding professors and colleagues at SMART for meaningful discussions and valuable feedback. I also thank helpful staff at SMART for their prompt administrative support. In addition, I am grateful to SMART for sponsoring my trip to Alaska to attend the IEEE Intelligent Transportation Systems Conference in September 2012. This gave me an opportunity to share and discuss my research interests with many researchers from academia and industry.

My sincere thanks go to Prof. Amedeo Odoni and Dr. George Kocur at the Department of Civil and Environmental Engineering (CEE) for their dedicated teaching. Their thought-provoking lectures helped me to broaden my knowledge, which is very essential for carrying out this research. I also gratefully acknowledge Prof. Nigel Wilson for his early help and guidance at the beginning of the MST program. Many thanks also go to other professors and administrative staff at CEE, particularly Prof. Cynthia Barnhart, Ms. Patricia Glidden and Ms. Kris Kipp for their constant support.

I am grateful to a fantastic group of professors, colleagues and friends at Laboratory for Information and Decision Systems (LIDS) and Aerospace Robotics and

Embedded Systems (ARES) group for sharing their interesting research topics and initiative ideas through numerous seminars, talks and weekly group meetings. I am also thankful to the MIT Transportation Club and Transportation@MIT for hosting the annual Transportation Showcase, which is an interesting event for me to present my work to enthusiastic audiences and to know state-of-the-practice in industry. I also owe thanks to many friends in the MST Class of 2013 for their warm friendship. Especially, I would like to acknowledge the generous financial support of the Office of the Dean for Graduate Education (ODGE) during my childbirth accommodation period. The support provided me with lots of flexibility and confidence to handle both personal life and research work at the same time.

My heartfelt gratitude goes to my mother, Truong Thi Minh Hanh, and my mother-in-law, Ho Thi Tuyet Ngoc, for sacrificing their working time to fly to the U.S., and for helping me to take care of my newborn baby during the first seven months of her life. This thesis could not have been written without their tremendous support and patience. To my father, Bui Thien Lam, and my father-in-law, Huynh Van Xuan, thank you for teaching me the importance of core values and high education, and for inspiring me to follow my passion. I would also like to thank my sister, Ngan, my brother, Phuoc, and my sister-in-law, Nhung, for their love and support.

Above all, my most special thanks go to my partner and also my colleague, Vu Anh Huynh, for his never-ending love, encouragement and sympathy. Words really cannot articulate how important your support is for me to finish this work. Thank you for always being by my side and taking care of every aspect of my life from the very early stage of this research. I have learned a lot from you through our fruitful discussions everyday. Needless to say, your incredible comments and suggestions contributed to the successful completion of this thesis. I would also like to thank our daughter, Mai (Misa) Huynh, for putting up with me. Thank you for traveling with me for my work, and for going to school with me to attend numerous lectures and exams when you were in my tummy without any complaints. You are my source of inspiration, and I look forward to watching you grow.

Contents

1	Introduction	13
1.1	Motivations	14
1.2	Problem Statement and Objectives	18
1.3	Contributions	18
1.4	Thesis Outline	19
2	Literature Review	21
2.1	The Fundamental Diagram of Traffic Flow	22
2.2	Traffic Assignment Models	25
2.2.1	Static Traffic Assignment Models	25
2.2.2	Dynamic Traffic Assignment Models	30
2.3	Congestion Pricing	33
2.3.1	Marginal Cost Pricing	34
2.3.2	Congestion Pricing Schemes	36
2.4	Game-theoretic Approaches	38
2.4.1	Game Theory and Mechanism Design	38
2.4.2	Algorithmic Mechanism Design	42
2.4.3	Vickrey-Clarke-Groves Mechanisms	43
2.5	Shortest Path Algorithms	47
2.5.1	Unconstrained Shortest Path Problems	47
2.5.2	Constrained Shortest Path Problems	48

3	Problem Formulation	51
3.1	Preliminary Notations and Definitions	51
3.1.1	Graph	51
3.1.2	Poisson Process	52
3.1.3	Mechanism	53
3.2	Assumptions	54
3.3	Dynamic Network Model	55
3.4	User-Centric Dynamic Pricing Mechanism	56
4	Approach	59
4.1	Allocation Rule	59
4.2	Payment Rule	63
5	UCDP Computation	67
5.1	Traffic Simulation	67
5.2	Minimum Travel Time Computation	68
5.3	Dynamic Allocation and Pricing Algorithm	69
6	Experiments	73
6.1	Parallel-link Network	73
6.2	General Network with a Bottleneck	78
7	Conclusion	81
7.1	Summary	81
7.2	Future Directions	83

List of Figures

2-1	Observation points in (k, u) , (k, q) and (q, u) diagrams.	23
2-2	Greenshields model.	24
2-3	Graphical explanation of marginal cost pricing.	35
2-4	Mechanism design.	39
3-1	Examples of graphs.	52
3-2	Illustration of Poisson arrivals in time.	53
3-3	User-Centric Dynamic Pricing mechanism.	56
6-1	Parallel-link network.	74
6-2	Performance of the UCDP mechanism when the demand is 970 v/h. .	76
6-3	Passengers' travel time and public transport usage when the demand is 1500 v/h.	77
6-4	Generalized cost and toll.	77
6-5	Performance of the UCDP mechanism in comparison with the SO per- formance.	78
6-6	General network.	79
6-7	Results for the UCDP mechanism on the general network.	80

List of Tables

6.1 Network specifications for Fig. 6-1. 74
6.2 Network specifications for Fig. 6-6 79

Chapter 1

Introduction

Road traffic congestion is a serious problem in contemporary metropolitan areas. According to U.S. Department of Transportation (DOT), traffic congestion costs the United States \$200 billion annually in wasted fuel and lost time [1]. Remarkably, DOT estimates that drivers in metropolitan areas spend one-quarter of total annual travel time in congested roads [1]. In the effort of reducing this significant amount of unpleasant and inefficient travel time, transportation researchers have developed many instruments and strategies aiming to evenly distribute vehicles on the networks. In this thesis, we focus on designing real-time traffic routing systems that use congestion pricing as an efficient method to alleviate travel delays in urban areas. Since transportation networks are complex large-scale systems, researchers are often concerned with the computational complexity of the problem [2]. In addition, the unpredictable behavior of drivers make it extremely difficult to require them to follow the routing systems. The efficiency of the systems as well as fairness among passengers also pose great challenges to researchers [3,4]. Therefore, in this work, we provide a novel game-theoretic approach to construct a dynamic congestion pricing mechanism that aims to address the above challenges.

1.1 Motivations

Intelligent and scalable traffic routing for large cities has been an active research area in recent years. Many route guidance systems have been used to assist drivers in path-choice decision making by simply computing the shortest path from a source to a destination, regardless of the changing conditions of roadways [5]. More advanced systems are proposed to recommend routes to users after performing some computations from a macroscopic point of view [6]. In these systems, the authors consider static flows of multiple homogeneous users on the networks and attempt to compute an *a priori* distribution of the users on the road networks. This static traffic assignment (STA) problem can be solved by using the Wardrop principle [7] or Karush-Kuhn-Tucker conditions to achieve a user equilibrium (UE) solution or a social optimal (SO) solution, respectively. At the UE state, each user selects his fastest route, while at the SO state, the total travel time of all drivers is minimized. Although SO is more preferable for road authorities, it encounters an unfairness issue as some users are assigned longer paths to allow for the efficiency of the entire system. In addition, STA models assume that traffic demand is constant over time, thus they are unable to represent the variation of traffic flows during short specific periods of a day such as peak hours [8]. This limitation along with the unfairness issue have raised a question of STA tractability in real-world deployments. To tackle this, Dynamic Traffic Assignment (DTA) models have been developed by many researchers to capture the peaked nature of travel demand and the time-varying traffic flows [9–13]. Approaches to DTA problems can be classified into two main groups: analytical approaches (mathematical programming, optimal control, variational inequality) and simulation-based approaches [2]. Although current DTA models can describe the reality more accurately than the static models, these DTA models are either mathematically intractable or heavily dependent on simulation, thus computationally inefficient when dealing with large-scale traffic networks. More importantly, they are still unable to manage the system optimality and the above unfairness problem simultaneously.

Thus, in [4] and [3], the authors propose the constrained system optimum (CSO) to capture the above two aspects. In their CSO models, they consider a traffic assignment pattern for homogeneous users that minimizes the total travel time subject to a fairness constraint via a *single* tolerance factor. However, this pre-specified tolerance factor does not truly reflect fairness for individual traveler whose preference and purpose are unrevealed. In addition, congestion can occur when the tolerance factor is not properly chosen even if the road network is capable of supporting congestion-free traffic at the SO state. Furthermore, the authors do not consider dynamic traffic flows and the selfish behavior of rational passengers. More precisely, a rational passenger would not use the routing system and follow its recommendation if such compliance does not provide him with a tangible gain. Apparently, these limitations have prevented a direct application of the model in real-world situations.

In fact, to partially restrain the selfish behavior of travelers that can lead to traffic congestion during rush hours, tolling systems such as Electronic Road Pricing (ERP) in Singapore have been deployed [14]. The positive effects of the ERP system on spreading peak-hour travel demand are verified in [15] and [16]. At the time of this writing, pre-computed time-varying tolls within a day are used to affect passengers' behavior. However, this time-varying tolling scheme does not enable congestion-free traffic flows because urban traffic also varies from day to day. A similar *variable price* scheme based on the evolutionary approach is proposed in [17] to enable congestion-free traffic when travel demand is static. In this scheme, passengers will decide to travel or stay at home permanently after a long adjustment process in response to variable tolls set by a road authority. Not only is the convergence process slow when the demand is static, but the traffic flow is also unstable when the demand is dynamic. As an attempt to compute flow-dependent tolls in dynamic environments, the work presented in [18] proposes a toll design method that combines road pricing with DTA. The authors formulate the problem as a bi-level programming problem, in which the road pricing model represents the upper-level decision making process, and the DTA model represents the lower-level counterpart. The resulting formulation is an NP-hard problem [18]. One possible method to obtain a locally optimal solution suggested by

the author is to combine grid search for the upper-level optimization and simulation for the DTA in an iterative procedure. However, such a method scales poorly with the size of the network and does not provide insightful understanding of the properties of obtained solutions for further improvement. We emphasize that all the mentioned pricing schemes have not addressed the unfairness issue among individuals.

Therefore, in recent years, many researchers have adopted a game-theoretic approach to handle the above difficulties from the microscopic perspective [19, 20]. Recent advances in algorithmic mechanism design [21] provide a promising approach to incentivize rational participants, or players, to cooperate with the system in order to reach desirable outcomes. This approach motivates players to disclose their private information. As usually done in market mechanism design [22], designers aim to construct a mechanism that has individual rationality (IR) and incentive compatibility (IC) properties. The former means that players do not suffer any loss when they use the system, and the latter means revealing truthful information is in their best interest. In algorithmic mechanism design, besides the IR and IC properties, designers also concern about computational complexity when computing an *allocation rule* and a *payment rule* for intended outcomes [23]. The key technical difficulties lie in the combinatorial nature of the allocation rule and the interweaving relationship of allocation rules and payment rules.

In our case, passengers can be viewed as players in a *routing game* created by the interaction among passengers and the road authority. Mechanism design for routing games have been studied extensively in computer networks [24–36] while very few works have been done for urban transportation networks [37–39]. Although both types of networks share some similar modeling aspects, inherent characteristics of the two types of networks differ substantially. In particular, in computer networks, as discussed in [40], utility functions of users are unknown to the network; and even if the utility functions are known, there is no central authority that knows all the link capacities and network topology. In addition, autonomous systems (AS) in the network can run any routing algorithm that benefits them the most. Therefore, we can hardly simulate flows for the entire computer networks such as the Internet.

In contrast, in transportation networks, passenger utility functions can be modeled quite accurately based on regression analysis [41]. In addition, we can monitor real-time traffic flows via sensor networks [42–45] or community based mobile navigation applications such as Waze [46]. Thus, current models and results in computer network routing cannot be applied in urban transportation network routing.

One of the most related works in mechanism design for urban transportation routing is [38]. In their work, the authors propose a *day-to-day auction mechanism* in which a road user has to bid everyday to purchase a bundle of network permits that allow him to use his preferred path. Although this approach can reduce the computational complexity of finding an allocation rule for a single OD pair network, it is not clear how to determine the allocation rule for a general network with multiple OD pairs. The day-to-day auction mechanism is also not practical to be implemented for real-world road networks due to several reasons. The mechanism requires users to go through a bidding process before being able to use the paths, while in reality, road users often needs an instantaneous route choice decision. Furthermore, although the authors show that truthfully reporting the valuation of several bundles of permits is a dominant strategy for each user, it is difficult for users to determine their true valuation.

To address the limitations in previous work, we explore a novel approach to overcome the computational complexity and to satisfy game-theoretic requirements of the transportation network routing problem. In this thesis, we propose a novel mechanism called User-Centric Dynamic Pricing (UCDP) based on recent advances in algorithmic mechanism design to allocate heterogeneous classes of users with paths that satisfy their travel preferences. Unlike other approaches that attempt to compute the path allocation for all drivers at the same time, the UCDP mechanism considers drivers in sequence. This idea helps to reduce the computational complexity of the problem. The mechanism is applicable, for example, in one-way car rental systems, or mobility-on-demand systems [47] where fleets of autonomous cars [48] are used in future cities to pick up and deliver passengers [49]. In the mechanism, these services are designed to work with public transportation systems.

1.2 Problem Statement and Objectives

We consider general road networks with dynamic flows of heterogeneous users. The UCDP mechanism is designed as follows. When a new passenger arrives at an origin node, he receives information from a road authority on the current network condition. The authority either suggests him to use public transportation if the network is about to be congested or provides him with the minimum travel time required to complete his trip. In the latter case, the passenger then reports his maximum tolerated travel time that is not less than the minimum travel time announced by the authority. Based on the current traffic condition in the network, the mechanism then offers the passenger a path to his destination.

Our objective is to assign each passenger a path that matches with his travel time constraint while maximizing sequential social surplus when this passenger joins the network. Importantly, we want to ensure that passengers have no incentive to lie about their preferences, and under our path assignment, congestion does not happen in the network. The UCDP mechanism takes into account the fairness concern when allocating paths to travelers. The tolls of paths are computed according to the marginal cost pricing principle.

1.3 Contributions

This thesis has four main contributions. The first two contributions describe important properties of the novel UCDP mechanism. The third contribution is the impact of the work to the transportation research community, and the last contribution presents interesting future work for practical deployments.

First, the mechanism is user-centric in the sense that it considers both passengers' preference and fairness among individuals. At the same time, the mechanism can achieve maximum sequential social surplus, i.e. the mechanism is efficient. We also prove that reporting true preference is a weakly dominant strategy of each passenger, and the mechanism never transfers positive amount of money to passengers.

Second, the mechanism introduces a new fairness constrained shortest path (FCSP) problem with a special structure that enables polynomial time computation of path allocation. Therefore, we can handle general networks having multiple OD pairs with dynamic flows of heterogeneous users in a computationally efficient way.

Third, to the best of our knowledge, this thesis is the first in the transportation literature that considers general road networks with dynamic flows of heterogeneous users and addresses system efficiency, fairness among passengers as well as computational complexity issues at the same time.

Fourth, this thesis lays a foundation for future work on designing distributed mechanisms for optimal traffic routing. In particular, we envision a large-scale network consisting of mobile users who not only share real-time traffic information but also collaborate with each other to achieve social-optimal traffic flows through our future distributed mechanisms.

1.4 Thesis Outline

This thesis is organized as follows:

- Chapter 2 provides an in-depth review of related works in road traffic routing and challenges arising from real-world deployments. Due to the multi-disciplinary nature of the problem, the review includes ideas and results from several fields ranging from transportation, economics to algorithm design. We first present traditional approaches such as static and dynamic traffic assignment as well as congestion pricing schemes in practice. We then discuss new game-theoretic approaches including mechanism design and algorithmic mechanism design, as they promisingly overcome limitations of traditional approaches. The discussion also highlights our motivation for adopting the algorithmic mechanism design approach in this thesis. Finally, we review VCG mechanism and important shortest path algorithms as a preparation step for our problem formulation and approach in the next chapters.

- Chapter 3 develops the formal problem formulation to the optimal traffic routing problem using our game-theoretic approach. We first provide preliminary notations, definitions and assumptions. We then formulate a model for a general network with dynamic traffic flows of heterogeneous travelers. We then focus on describing our User-Centric Dynamic Pricing (UCDP) Mechanism in detail.
- Chapter 4 presents our approach to design an allocation rule and a payment rule, which are two important components in our mechanism. The allocation rule assigns a path to each passenger and is formulated as a constrained shortest path problem. The payment rule calculates the toll that each passenger pays to use his assigned path. Most importantly, this chapter presents two theorems with proofs showing that the mechanism has incentive compatibility and non-positive transfer properties.
- Chapter 5 develops algorithms that are used to allocate paths and compute tolls in real-time. First, we show how to simulate a network state at each time instant. We then present an algorithm to determine minimum travel times announced by a road authority. Subsequently, we describe an algorithm for computing dynamic path allocation and dynamic tolls.
- Chapter 6 demonstrates the performance of the proposed mechanism through two simulated routing experiments. In the first experiment, we simulate the UCDP mechanism on a parallel-link network with one OD pair. We then compare the network performance under the UCDP mechanism with user-equilibrium and social-optimal performance. In the second experiment, we consider a general network with multiple OD pairs. We show how the road authority uses the UCDP mechanism to manage traffic flows when road conditions change over time.
- Chapter 7 concludes the thesis by summarizing the motivation and objectives of this thesis, the research conducted in this thesis as well as the key contributions. This chapter also discusses several interesting directions for future research.

Chapter 2

Literature Review

In this section, we extend the review in Section 1.1 to provide a thorough literature review of road traffic routing and challenges in real-world deployments. Due to the multi-disciplinary nature of the problem, this chapter presents important results in related fields ranging from transportation, economics to algorithm design. In particular, Section 2.1 first reviews the fundamental diagram that describes characteristics of traffic flows at the macroscopic level. A detailed description of Greenshields model for the fundamental diagram will also be included in this section, since it will be applied in our UC DP mechanism. Subsequently, Section 2.2 presents static and dynamic traffic assignment models, which have been widely used to understand the phenomenon of congestion and its causes. We then introduce the idea of congestion pricing in Section 2.3, which is based on the underlying economics of transportation systems to alleviate congestion. We also present several congestion pricing schemes in practice and relevant research. To further understand the selfish behavior of drivers, and the interaction between drivers and the road authority from a game-theoretic point of view, a brief review of game theory is then provided in Section 2.4. In particular, this section focuses on reviewing mechanism design, a sub-field of game theory, and its applications in previous works. In addition, a more advanced approach called algorithmic mechanism design is presented to address computational tractability and scalability arising from the mechanism design approach. For better understanding of

our payment rule explained in Chapter 4, a detailed description of Vickrey-Clarke-Groves (VCG) mechanism is also included in this section. Finally, Section 2.5 reviews important shortest path algorithms that are the building blocks for computing the allocation rule and payment rule in our proposed mechanism in Chapter 5. The set of rich results presented here will be incorporated in our novel game-theoretic approach to tackle the optimal dynamic routing problem.

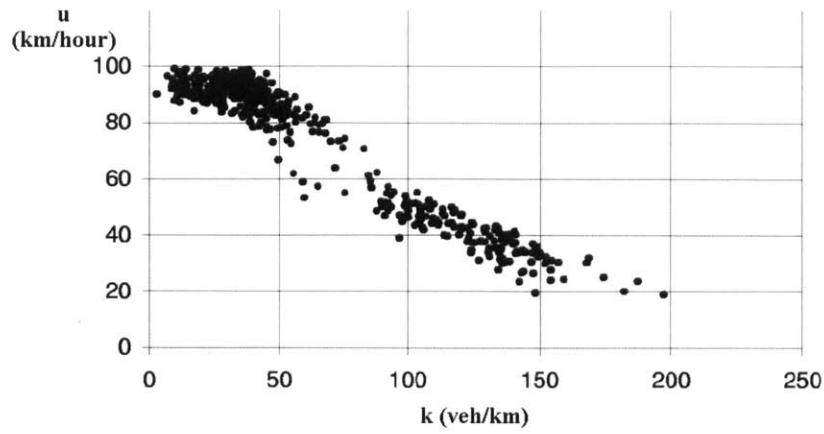
2.1 The Fundamental Diagram of Traffic Flow

The fundamental diagram characterizes the relationships among three macroscopic variables of traffic flow on a road segment including average density k , average flow q and average speed u . The average density k reflects the number of vehicles per length unit of the road, and average flow q represents the number of vehicles that passes a certain point per unit of time [50]. Under static conditions such as long roads and long time periods, the average speed u can be calculated based on the fundamental relation of traffic flow theory as follows:

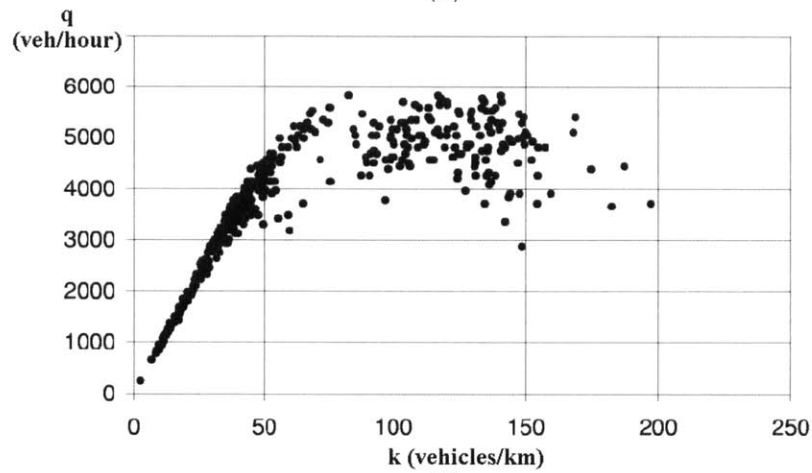
$$q = uk. \tag{2.1}$$

The (k, q) curve that expresses the relationship between flow and traffic density is called the fundamental diagram, since it represents all the three variables. The fundamental diagram can be derived by plotting field data points and finding a best fit curve for these data points. For example, Immers et al. observed and measured flow and mean speed on a three-lane motorway during one minute [50]. Their observation points are illustrated in Fig. 2-1.

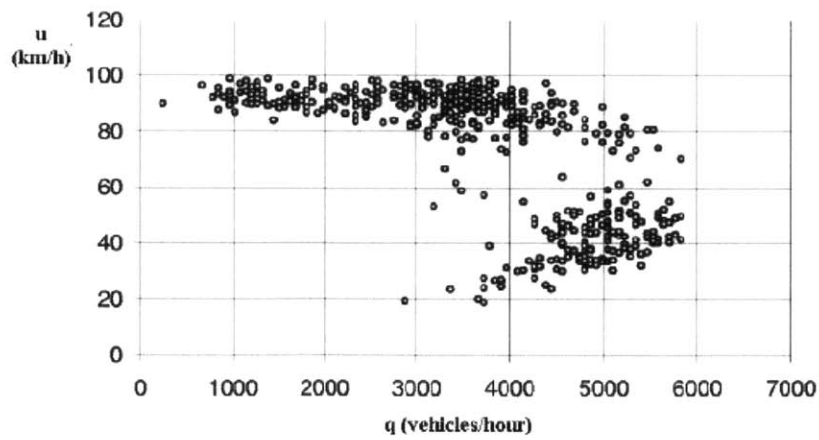
Research on the fundamental diagram began in 1933, when Greenshields calibrated traffic flow, traffic density and speed by using a photographic measurement method and then derived a mathematical model for the fundamental diagram [51, 52]. As shown in Fig. 2-2(a), in the Greenshields model, he postulated a linear relationship between speed and traffic density, in which speed declines linearly with traffic density,



(a)



(b)



(c)

Figure 2-1: (a) Observation points in (k, u) diagram, (b) Observation points in (k, q) diagram, (c) Observation points in (q, u) diagram [50].

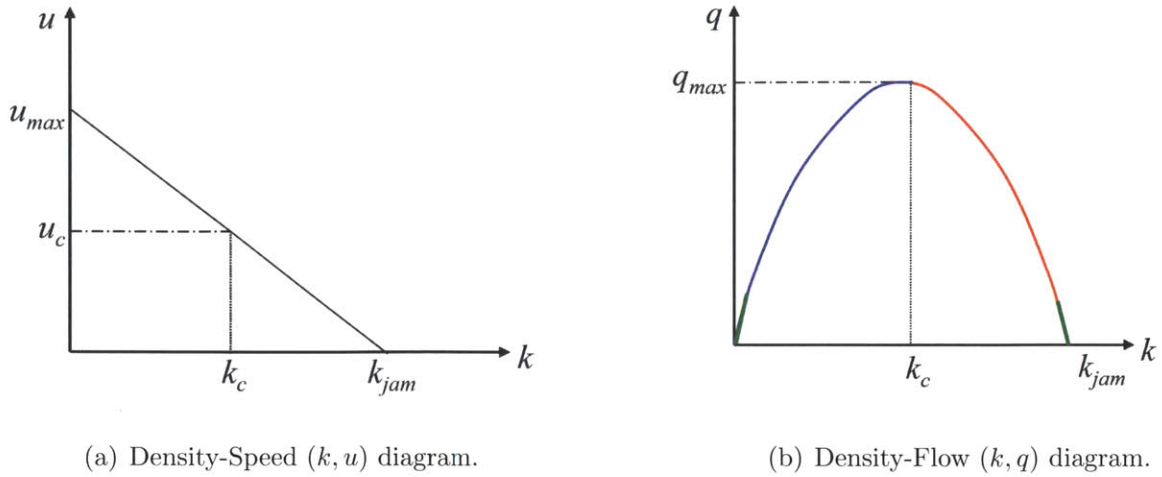


Figure 2-2: Greenshields model.

taking its maximum value u_{max} when $k = 0$ and declining to $u = 0$ when $k = k_{jam}$, the jam density. Thus, the basic relationship for the Greenshields model is:

$$u = u_{max} \left(1 - \frac{k}{k_{jam}} \right). \quad (2.2)$$

The linear relationship between speed u and traffic density k leads to a parabolic relation for the (k, q) diagram shown in Fig. 2-2(b). Combining Eq. 2.1 and 2.2, we can write:

$$q = u_{max} k \left(1 - \frac{k}{k_{jam}} \right). \quad (2.3)$$

In the (k, q) diagram, the value k_c for which flow q reaches its maximum q_{max} is called critical density, since it marks the start of an unstable flow area where additional input of vehicles on a road segment decreases traffic flow and eventually leads to traffic congestion, i.e., $q = 0$ and $k = k_{jam}$. By differentiating the right hand side of Eq. 2.3 and setting the derivative equal to zero, we obtain:

$$k_c = \frac{k_{jam}}{2}. \quad (2.4)$$

Substituting Eq. 2.4 into 2.2, we also find that the value u_c associated with k_c and

q_{max} is equal to $\frac{u_{max}}{2}$. Replacing k with $k_c = \frac{k_{jam}}{2}$ also provides:

$$q_{max} = \frac{u_{max} \cdot k_{jam}}{4}. \quad (2.5)$$

Since the Greenshields model requires specifications of only two parameters u_{max} and k_{jam} , it is especially attractive for practice use. In this thesis, we use the Greenshields model to estimate and to predict traffic flow in our experiments for the proposed dynamic pricing mechanism. In a real world deployment of our system, we can obtain more accurate real-time traffic flow information via sensor networks.

2.2 Traffic Assignment Models

2.2.1 Static Traffic Assignment Models

With the rapid growth of automobiles industry at the beginning of 19th century, road congestion had soon become a serious problem in modern cities and has captured strong interest of many researchers. To understand drivers' behavior and traffic patterns that are subject to congestion, in 1952, Wardrop conducted several experiments to figure out the relation of speed and flow on a given road segment in one direction [7]. As a result, he derived two sound principles that define the two well-known notations of traffic equilibria. In particular, according to Wardrop's first principle, the network is considered stable when the travel times in all used routes are equal and less than those which would be experienced by a single driver on any unused route. In other words, given that the travel time on every route on the network are publicly known, each driver non-cooperatively chooses his best route and has no incentive to switch to other routes since he can not further reduce his travel time. The traffic flows that have this property are defined as User Equilibrium (UE) flows. Different from the first principle, Wardrop's second principle assumes that drivers consistently make route choice decisions cooperatively with each other to minimize total system travel time. We can also imagine that there exists a central authority who tell users which

routes to follow, and they fully comply with these suggestions to maximize the system performance. Under this behavior, when the average travel time is minimized, the second principle claims that the network reaches its equilibrium state called Social Optimal (SO) state.

Over the last five decades, Wardrop’s principles have inspired many transportation specialists to formulate and solve mathematical models to obtain the UE and SO flows in urban transportation networks [6, 53–59]. Their works are often classified as static traffic assignment (STA) problem, which is described in [60] as follows. Given the inputs: i) a graph representation of an urban transportation network, ii) the associated link performance functions, and iii) an origin-destination matrix that shows the trip rates between all OD pairs during the period of analysis, the STA problem is finding the traffic flow on each link on the network. The link travel time can then be computed by using link performance functions. The STA problem aims to help transportation planners and road authorities to: i) estimate the distribution of traffic flows and travel times on all links given a level of demand, ii) identify heavily congested links, and iii) make appropriate physical changes to the road network such as adding one or more links. The last goal is extremely important because adding more resources to the network might deteriorate the network performance, i.e., increase delays and congestion instead of decreasing them. This counter-intuitive phenomenon is known as Braess paradox [61] and has been studied extensively in [62, 63].

The first mathematical model to solve the above STA problem was proposed in 1956 by Beckmann et al. [53]. This pioneering work of Beckmann has laid a foundation for a variety of extended works in the field (see [6, 57, 58] and references therein). Sheffi [60] shows that Beckmann’s User-Equilibrium model can be expressed as a nonlinear mathematical optimization problem:

$$\min z(\mathbf{x}) = \sum_a \int_0^{x_a} t_a(\omega) d\omega$$

subject to

$$\begin{aligned} \sum_k f_k^{rs} &= q_{rs}, \quad \forall r, s, \\ x_a &= \sum_r \sum_s \sum_k f_k^{rs} \delta_{a,k}^{rs}, \quad \forall a, \\ f_k^{rs} &\geq 0, \quad \forall k, r, s, \end{aligned}$$

where k is a path that connects origin r and destination s , x_a is the flow on link a , t_a is the travel time on link a , f_k^{rs} is the flow on path k connecting OD pair (r, s) , q_{rs} is the trip rate between r and s . The indicator variable $\delta_{a,k}^{rs}$ is equal to 1 if link a is on path k between OD pair (r, s) and 0 otherwise.

The solution for the above formulation is the link flows \mathbf{x} that satisfy the UE conditions and reflect that all the trip rates q_{rs} have been distributed appropriately. The objective function is the sum of the integrals of the link performance functions. At the first glance, the objective function looks obscure. In fact, this is a potential function of a routing game among drivers. We will comment on this potential game formulation later when we review game theory and mechanism design in Section 2.4.1. The first two constraints are flow conservation constraints, which guarantee that all trip rates have to be assigned to the network. The last constraint is nonnegativity constraint, which is used to ensure that the obtained solution will be physically meaningful. Beckmann showed that this mathematical programming formulation is equivalent to the UE assignment problem. He also proved the existence and uniqueness of the solution. In addition, it is shown that the above formulation is a convex optimization program since the link travel time function is a monotonically increasing function.

Beginning in the late-1960s and continuing into the mid-1980s, many researchers had proposed different approaches to solve Beckmann's UE formulation including heuristic and convex combination approaches. The two most common heuristic methods are *capacity restraint* and *incremental assignment*, which were originally proposed by Mosher in 1963 [54] and Martin and Manheim in 1965 [55] respectively. Although

these techniques are embedded in many early transportation planning computer packages [64, 65], they were criticized due to the lack of convergence properties [66, 67]. The capacity restraint methods are not guaranteed to converge, and incremental assignment methods may converge to a nonequilibrium solution. Apparently, they can not obtain a set of flows that satisfy UE conditions. This unsatisfactory performance of heuristic approaches has posted a strong need for a novel numerical optimization method. Fortunately, a novel method called convex combination algorithm proposed by Frank and Wolfe in 1956 [68] is capable of solving quadratic programming problems with linear constraints. Not until thirteen years later, in 1969, Bruynooghe discovered Frank and Wolfe' work and suggested an application of convex combination method to solve the transportation network equilibrium problem [56]. Basically, in this method, flows are taken from more congested paths and assigned to less congested paths until flow changes are small. The method then uses a shortest path algorithm to solve a linear program subproblem because the algorithm is very efficient even for large networks. The convex combination method has been proven through numerous research to be a simple and efficient method to the minimization of the UE program [69–72]. Although the method converges slowly, it is still popular since other faster methods require more memory and do not use shortest path subproblem naturally.

Regarding the SO flow pattern, Beckmann's SO formulation is described in [60] as follows:

$$\min z(\mathbf{x}) = \sum_a x_a t_a(x_a)$$

subject to

$$\begin{aligned} \sum_k f_k^{rs} &= q_{rs}, \quad \forall r, s, \\ x_a &= \sum_r \sum_s \sum_k f_k^{rs} \delta_{a,k}^{rs}, \quad \forall a, \\ f_k^{rs} &\geq 0, \quad \forall k, r, s. \end{aligned}$$

The objective function of the SO formulation is more straightforward than in the UE formulation. It aims to minimize total travel time spent in the network subject to the same constraints as in the UE equivalent program. However, as pointed out in [60], the SO formulation does not generally generate an equilibrium flow pattern since some travelers can be better off by unilaterally switch routes. Hence, the SO flow pattern is not stable and should not be used to represent the actual behavior of drivers and equilibrium state. It can be used to observe the best-case scenario when all drivers cooperately make decisions. The SO formulation is also helpful when we want to compare different static traffic assignment models [73]. In particular, we can use a popular measure named the *price of anarchy* [74], which is the ratio between the total utility at UE and SO state to see how far the UE performance is from the best possible use of the network. The SO formulation can be solved by using the first-order necessary conditions of the associated Lagrangian function. Section 3.4 in [60] provides the details of the method. We note that the solution to the SO formulation can also be found by using the convex combination method mentioned above.

Parallel to the development of algorithms to solve Beckmann’s formulation, extensions of Beckmann’s model have also been investigated (see [59] and references therein). These extended models often introduce side constraints to become more realistic than the traditional traffic assignment models. For example, the constraints can be used to model the effects of a traffic control policy or to describe the link capacity restrictions set by a road authority. However, these extended models have not been fully developed and solved due to associated computational issues.

Despite of all of the hard work to craft and solve existing STA formulations, the STA models’ tractability in real-world deployments is still a major concern since they have some fundamental issues. First, because the UE flow pattern is rooted in drivers’ selfish behavior, attaining the UE flow pattern does not help in preventing or alleviating traffic congestion. Therefore, the SO flow pattern, which results in better network performance, is more preferable for a central authority. However, as some drivers suffer longer paths than others to allow for the efficiency of the entire system, and they naturally want to change to other shorter paths, the SO flow pattern is

unfair and unstable. Second, the STA models assume that OD flows are constant within the considered period in order to apply steady-state analyses. To keep this assumption accurate, the period of analysis can only be very small. However, this requirement is not reasonable since the STA models are only meaningful when their period of analysis is longer than the typical duration of trips at this time. Therefore, STA models are not sufficient to describe the dynamics of traffic flows during a very short specific period of time such as peak hours, when flows are changing rapidly and unpredictably.

2.2.2 Dynamic Traffic Assignment Models

The limitations of STA models have motivated researchers to develop Dynamic Traffic Assignment (DTA) models to adequately represent traffic reality and drivers' behavior. Departing from STA to deal with time-varying traffic flows, DTA refers to a wide range of problems, each depends on different sets of decision variables, system inputs, assumptions about systems and users, and modeling purposes. According to Peeta and Ziliaskopoulos [2], approaches to DTA can be classified into two main groups: analytical approaches (mathematical programming, optimal control, variational inequality) and simulation-based approaches.

Analytical approaches:

In 1978, Merchant and Nemhauser (hereafter referred to as the M&N) proposed the first DTA model, which is a discrete time, nonlinear and nonconvex mathematical programming problem [9, 10]. The M&N model only considers deterministic, fixed-demand, one destination, one commodity and SO cases. Basically, there are a link exit function that aims to propagate traffic and a static link performance function that computes travel cost based on link volume. The global optimum of the M&N formulation can be obtained by solving a piecewise linear version of the model with additional assumptions on the objective function by using a one-pass simplex algorithm. Later, Ho suggested a stepwise linear version of this model and pointed out that we can also achieve the global optimum by solving a sequence of at most $N+1$

linear programs, where N is the number of periods [75]. Following Ho, Carey showed that the M&N problem can be reformulated as a well-behaved convex nonlinear program and solved by standard mathematical programming software [76]. Carey also extended his models to handle multiple destinations and multiple commodities case. Unfortunately, the resulting formulation does not satisfy many requirements of general networks such as “first in first out” (FIFO) and no “holding back” (“holding back” is an issue arising when traffic at certain approaches is held back to let other traffic streams move faster to minimize system total delays, and this creates unfair or unreasonable SO flow pattern). To address the above unsatisfaction, substantial research on DTA using mathematical programming with different sets of assumptions have been conducted [77–79]. Nevertheless, attaining both mathematical tractability and high fidelity of real traffic flows is still an open problem.

Another approach to DTA is known as optimal control formulation, in which the OD trip rates and link flows are continuous functions of time. The constraints are similar to those in mathematical programming formulation but are modified to adapt in a continuous-time setting. The initial link-based UE and SO optimal control formulations for one destination case are suggested by Friesz et al. [80]. Following Friesz, Wie extended the UE model to capture the elastic time-varying travel demand [81]. Ran and Shimazaki then formulated a link-based SO model for general networks [82]. However, this model overlooked the FIFO issue and can only be applied for a very small scale network. Subsequently, various optimal control theory-based DTA models are discussed in the literature [83–85], but they are based on unrealistic assumptions and lack of additional constraints to manage FIFO and “holding back” issues. Importantly, efficient algorithms for solving these formulations are still questionable.

To circumvent the problems associated with optimal control theory-based models, another approach called variational inequality (VI) has been researched. VI problem, as defined by Nagurney in [57], is a general problem formulation that includes a plethora of mathematical problems such as nonlinear equations, optimization problems, complementarity problems, and fixed point problems. Therefore, VI formulation is applicable for traffic assignment problem since it allows for a unified treatment of

equilibrium problems and optimization problems. Originally, this approach was introduced by Dafermos in 1980 for solving the STA problem [86]. A decade later, a continuous time VI model for departure time and route choice decision making was suggested by Friesz et al., but no proof of solution existence or uniqueness as well as no efficient algorithm were provided [87]. Wie et al. then formulated the problem as a discretized VI model, in which they replaced exit time functions used in Friesz’s formulation by exit flow functions [88]. They also devised a heuristic algorithm to solve it approximately. Although the existence of a solution under some certain conditions was proved in this work, the need for an efficient algorithm still remains since this path-based formulation requires tremendous computational efforts to enumerate complete paths. To reduce the computational burden of path-based models, several link-based models were proposed [89–91]. While these VI models are more general and provide greater analytical flexibility to address multiple DTA problems than other analytical models, it turns out that they expose to more severe computational feasibility issues. Furthermore, they are still unable to overcome challenges arising in realistic networks such as FIFO and “holding back”. As a result, in recent years, DTA models have migrated toward new approaches called simulation-based approaches, discussed hereafter.

Simulation-based approaches:

The goals of simulation-based DTA models are to use a traffic simulator to replicate dynamic traffic phenomena and to search for the optimal traffic flow propagation. More crucially, researchers aim to use the simulator iteratively to forecast the future traffic conditions. Thanks to this promisingly descriptive and predictive power, the transportation literature has witnessed a boom in research on simulation-based DTA models since the first well-recognized work of Van Aerde and Yagar in 1988 [92,93]. Traffic simulators have different types depending on the choice of granularity such as macroscopic, microscopic or mesoscopic. At first, several microscopic models have been developed for small scale networks, and some are actually embedded in commercial software such as AIMSUN2 [94] and INTEGRATION [92,93]. Later, mesoscopic models, which are a combination of a microscopic representation of indi-

vidual vehicles and macroscopic description of interactions among traffic flows, have been evolved to cope with large-scale networks. Some recognized mesoscopic models are DYNASMART [95, 96], and DYNAMIT [97]. Although these models circumvent some persistent modeling issues in analytical approaches by directly addressing them through simulation process, it is impossible to derive proofs of existence, uniqueness and convergence of solutions due to the lack of theoretical formulations. In addition, the computational cost associated with the use of a simulator is still too high to make the models applicable in practice. The underlying difficulty lies in selfish behavior of drivers. As drivers independently and constantly make route choice decisions that are in their best interest, predicting future traffic flows requires lots of memory and long computation time. Therefore, recently, there has been a heightened interest in searching for deployable traffic policies to positively affect drivers' behavior, and incorporating the policies into models to reliably represent and predict real traffic maneuvers. The next section reviews congestion pricing as one of these policies and its application in practice.

2.3 Congestion Pricing

As urban transportation exhibits the phenomenon of diseconomies of scale, the idea of congestion pricing is based on the underlying economics of transportation systems. This approach was initiated by Pigou in his famous book *The Economics of Welfare* (1920) [98]. Pigou used an example of a congested road to show that travelers were not using roads efficiently because they did not have to pay for the congestion costs they imposed on others. A price mechanism that charges suitable fees on different roads would motivate travelers to use the facilities more efficiently and rebalance traffic flows. This would reduce aggregate travel time and thus increase social benefits. Since Pigou's first work, congestion pricing has received remarkable attention in the scientific literature for decades (refer [99] and references therein). In the following discussion, we will review the marginal cost pricing theory and practical congestion pricing schemes.

2.3.1 Marginal Cost Pricing

Theoretically, research in congestion pricing has relied upon the fundamental economic principle of *marginal cost (first-best) pricing*. The marginal cost pricing principle was developed based on demand-supply curves for the standard case of a traffic flow with homogeneous users moving inside a given network [100–102]. The idea is that each driver should pay a cost or congestion charge that is equal to the external congestion costs that he imposes on all other drivers. This marginal cost is the difference between the cost of road usage to an individual and the social cost of adding one extra vehicle to the traffic stream. Formally, the marginal cost pricing principle was summarized in [103] as follows.

Let q be the traffic flow, in terms of the number of vehicular trips per unit of time, w be the level of service or capacity (this is commonly assumed to be fixed), and $v(q, w)$ be the average cost per vehicular trip or marginal private cost (MPC) that a road user spends when operating a vehicle on road facility. Then, the total opportunity cost of congestion is:

$$V(q, w) = q \times v(q, w). \quad (2.6)$$

The optimal road price p or marginal social cost (MSC) of an extra vehicular trip can be derived by taking the first derivative of Eq. 2.6 with respect to q :

$$p = \frac{\partial V}{\partial q} = v(q, w) + q \frac{\partial v(q, w)}{\partial q}. \quad (2.7)$$

Eq. 2.7 shows that the optimal road price p has two components: the marginal private cost, plus the change in the marginal private cost from serving an additional user. According to the marginal cost pricing principle, the optimal congestion charge τ that each driver should pay is exactly equal to the second component of Eq. 2.7:

$$\tau = q \frac{\partial v(q, w)}{\partial q}. \quad (2.8)$$

The concept of marginal cost pricing is intuitively illustrated in Fig. 2-3, which depicts the curves of the MSC function, given by $\partial V / \partial q$, the MPC function $v(q)$ (assuming that w is fixed) and a demand function. The optimal congestion charge τ imposing on each driver is equal to the difference between two points: the intersection of the MSC curve and the demand curve, and the MPC at social optimal flow q_{so} level. In other words, τ closes the gap between the MPC and the optimal road price p for facility use, shifts traffic from user equilibrium state q_{ue} to q_{so} and creates a net welfare benefit (patterned area). Given fixed road capacity, an increase of the travel demand would lead to higher congestion charge and a larger net welfare benefit.

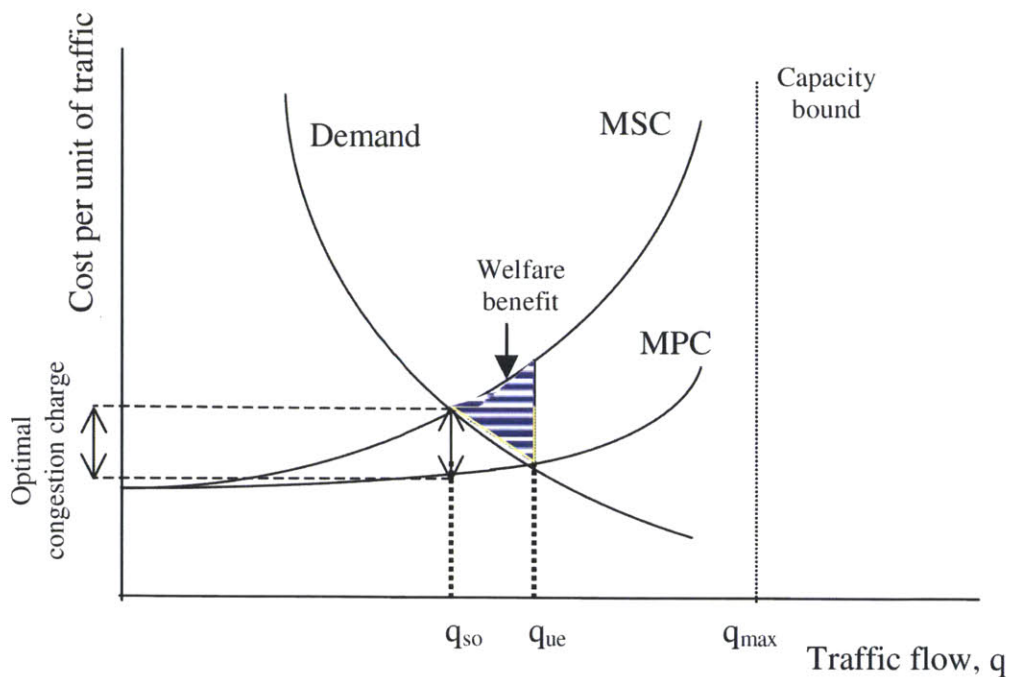


Figure 2-3: Graphical explanation of marginal cost pricing: the optimal congestion charge imposing on each driver helps to shift the traffic from the user equilibrium flow level q_{ue} to the social optimal flow level q_{so} and yields a net welfare benefit [103].

2.3.2 Congestion Pricing Schemes

Although the marginal cost pricing theory is appealing, this pricing scheme has not been implemented in practice due to technological difficulties arising when computing a payment for every traveler. The first congestion pricing scheme in practice is Area Licensing Scheme (ALS), which was launched by Singapore in 1975 [104, 105]. The ALS system charged drivers who crossed a cordon line around the central business district (CBD) during peak hours on week-day mornings about US\$1 a day or \$20 a month. Although the ALS reduced 75 percent of private vehicle travel and 50 percent of all vehicles entering the CBD, surprisingly the actual travel times per bus or auto trip to a destination in CBD did not change due to the disruptive effects of traffic rerouting to the free of charge peripheral roads. The high charges also yielded under-utilization of the road network and shifted congestion to expressways and non-restricted time. Thus, in 1995, a paper-based Road Pricing Scheme (RPS), operating in the same way as the ALS, was introduced on an expressway (East Coast Parkway) and later extended to other expressways to further alleviate congestion [105]. One problem associated with both systems was limited temporal and spatial variations in charges due to the lack of automation. In fact, paper-based schemes required personnel to operate the system, and this led to high probability of mistakes by operating officers. In addition, there was always a rush to enter just before or after the restricted hours that could not be smoothed by a shoulder-peak charge. To handle this problem, in 1998, the government introduced a new system namely Electronic Road Pricing (ERP) to replace the current system [105]. Gantries were installed at all the approach roads to the CBD zone and on the expressways. Different from the ALS, the ERP scheme charges vehicles each time they pass through a gantry. The fees are automatically deducted from CashCards, a smart card inserted in the In-vehicle Units (IUs) of vehicles. The current ERP scheme has different charges according to vehicle type, time of the day and gantry locations.

Following Singapore, a variety of congestion pricing schemes have been deployed worldwide such as High Occupancy Toll (HOT) and High Occupancy Vehicle (HOV)

lane facilities in California and the Twin Cities in the U.S., congestion charge facilities in London, Stockholm, Valletta and Milan. More examples and details can be found in [103, 104, 106]. To increase the effectiveness of congestion management, several systems vary tolls according to a predetermined schedule. For instance, on State Route 91 (SR91), the primary link between Orange and Riverside counties south and east of Los Angeles, as of July 2013, tolls vary between \$1.40 in the early morning hours and \$9.55 during peak hours [107]. In Singapore, the ERP rates are adjusted regularly depending on time of day, day of week (weekday vs weekend) and season. However, these pre-computed time-varying tolling schemes do not fully prevent congestion, since urban traffic also varies continuously within a day and across days.

As an attempt to compute flow-dependent tolls in dynamic environments, Joksimovic proposed a toll design model that combines road pricing with DTA by using bi-level mathematical programming [18]. In particular, a bi-level program contains the upper level, where a road authority decides tolls on links, and the lower level, where travelers independently respond to the tolls by changing their travel behavior. More precisely, travelers try to maximize their utilities to reach the UE state for the announced toll level, while the road authority aims to optimize his objective given that the traffic flows would be at the UE state. In Joksimovic's model, the classical DTA model is extended and modified to capture impacts of road pricing on drivers' decision and serves as the lower level of the problem, while the road pricing model serves as the upper level of the problem. The complexity of the problem is NP-hard, as pointed out by the author. The author then suggested a two-stage iterative procedure for determining the optimal tolls that the road authority should charge on links in order to reach his objective. However, such iterative solution approach is not scalable and does not provide insightful understanding of the properties of obtained solutions for further improvement. For example, one cannot infer the behavior of an individual traveler if the road authority changes the toll level slightly or the influence of travelers on each other. Therefore, recently, many researchers have adopted game-theoretic approaches to understand the problem from the microscopic perspective [108–110].

We are going to overview this approach in the next section.

2.4 Game-theoretic Approaches

2.4.1 Game Theory and Mechanism Design

Game theory is a mathematical way to describe strategic reasoning of rational decision-makers. Since 1950, game theory has been widely studied in several disciplines such as economics, political science and biology to understand the competition and cooperation among agents and the role of threats or punishments in long term relations. By definition, a game must consist of multiple players, each player must make a decision, and each player has a *utility function* that depends on his own decision and other players' decisions [111]. In a game, all players independently maximize their utility functions. As traffic routing problem involves multiple travelers as rational players, and each player needs to make his traveling decision to optimize his utility function, the problem can be seen as a non-cooperative or selfish *routing game*. Therefore, several results from game theory can be used to analyze and characterize properties of the problem.

Although previous works in traffic routing have not directly investigated the problem from the game-theoretic perspective, some aspects of game theory has been touched slightly. For example, under game theory terminologies, it turns out that Wardrop's first principle on the UE traffic flow pattern coincides with Nash equilibrium solution concept of a game [111]. For another example, the objective function in Beckmann's UE formulation discussed in Section 2.2.1 is in fact a potential function of the routing game, which is a potential game. Potential games are games that admit a potential function as a function of all players' decisions that gives information about each player's utility function. In particular, the maximization of a potential function with respect to a player's decision coincides with the maximization problem of that player [111]. Similarly, when congestion pricing is introduced for traffic flow management, the mentioned bi-level programming problem is a single leader rest follower

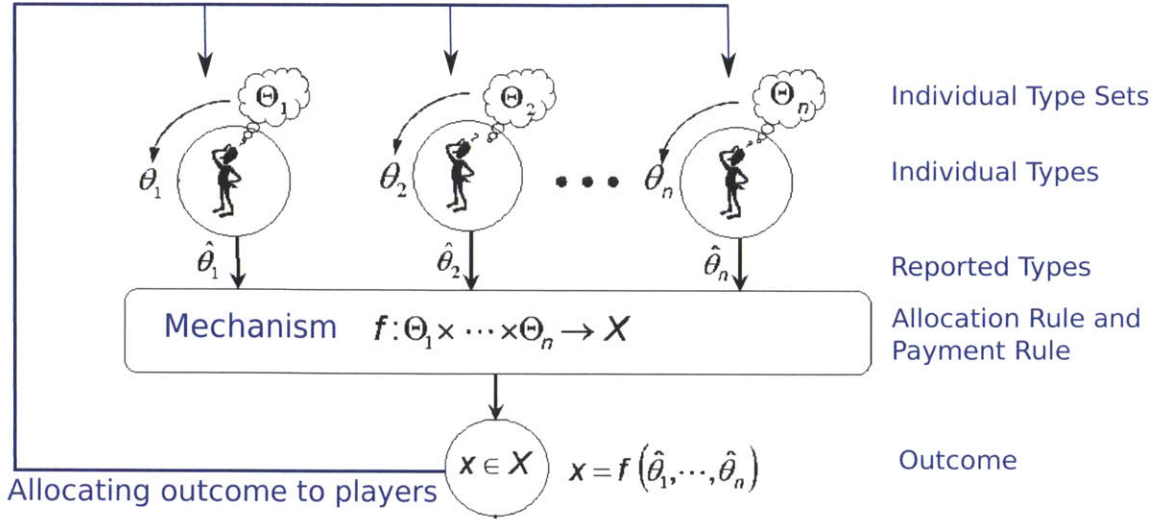


Figure 2-4: Mechanism design.

problem and is known as a *Stackelberg game* [110].

The above induced games are formed naturally by selfish behavior of travelers and the interaction among travelers and the road authority. We can use results from the game theory literature to further analyze the existing games. However, it is well-known that computing Nash equilibria is computationally hard in a general game [112]. Therefore, a new trend in game theory is to design games with special structure that are easy to be solved for equilibria. This sub-discipline is called *mechanism design* and has been studied intensively by economists to construct market mechanisms and auction mechanisms [113].

The mechanism design problem is to allocate resources among players in a strategic setting, assuming that each player acts rationally to maximize his utility. The goal is to aggregate individual preferences into a desirable outcome or a collective decision. A mechanism is depicted in Fig. 2-4. Given that there are n players, each player has a type θ_i , which is his private information or preference. Let Θ_i denote the set of possible types of player i and $\Theta = \Theta_1 \times \dots \times \Theta_n$ denote the set of all possible types of all players. Each player then reports his type called reported type $\hat{\theta}_i$ to the mechanism. Subsequently, the mechanism aggregates all players' reported types into a desirable outcome x by applying function $f: \Theta \rightarrow X$, where X is the set of all

possible outcomes. Finally, the mechanism allocates resources to players based on the computed outcome x .

An important feature of the mechanism design problem is that individuals' actual preferences are not publicly observable. Incentivizing agents to reveal their private information would significantly simplify the problem to achieve this goal. When being asked, an agent can lie about his true information. If a mechanism ensures that each agent's best interest is to truthfully reveal his private information regardless of what the others do, we say that the mechanism is *incentive compatible* (IC) or *strategy-proof* [21]. We, however, can not force an agent to join a game created by a mechanism which offers him less expected utility in comparison to not joining the game. Thus, we call a mechanism is individual rational (IR) if for all agents, it does not incur any loss to join the resulting game. Constructing a mechanism with both IC and IR properties is desirable in mechanism design. Besides IC and IR, mechanism designers also concern about computational complexity when computing an *allocation rule* for achieving intended results and a *payment rule* that motivates the agents [23]. The key technical difficulties lie in the combinatorial nature of the allocation rule and the interweaving relationship of allocation rules and payment rules.

The simplest method to obtain agents' private information is to ask them directly. This type of mechanisms is called direct mechanisms. Indirect mechanisms would elicit a function of private information from agents. Without loss of generality, we can focus on incentive-compatible direct mechanisms based on a well-known result named *Revelation Principle*. This principle proves that if there exists an equilibrium with dominant strategies in an indirect mechanism, then we can construct an equivalent incentive-compatible direct mechanism, in which agents truthfully report their private information and are allocated services and charged prices accordingly [21, 114].

Influenced by the use of mechanism design in auction theory context, several transportation researchers have come up with different tradable permit schemes, in which the road space is treated as a common commodity. Teodorovic et al. proposed an auction-based congestion pricing scheme to reduce traffic congestion in a downtown area [37]. Under this scheme, all drivers who want to access to a cordoned downtown

area in a specific time period must participate in an auction and submit bids to a road authority acting as an auctioneer. The bids vary according to the duration of drivers' visit to the downtown area. The road authority then can decide whether to accept or reject particular bids by the drivers. Later, a similar scheme called tradable bottleneck permits was suggested by Akamatsu [115] and then was extended in [38]. This day-to-day auction mechanism requires a traveler to bid everyday to purchase from the road authority a bundle of network permits that allow the traveler to use his preferred path. Although the authors show that truthfully reporting the valuation of several bundles of permits is a dominant strategy for each user, it is difficult for users to determine their true valuation. Since each permit is only valid for a link on a specific time, obtaining sufficient permits for the whole path requires a large number of transactions, especially when the network is complex. To simplify the system, Yang et al. investigated an alternative tradable credit scheme in which travel credits hold by drivers can be used on any link, but each link has different amount of credit charge [39]. At the beginning, government will distribute free credits to travelers and then travelers can buy or sell their credits in a free credit trading market without the government's interference.

While tradable permit schemes are actually deployed in practice under Kyoto Protocol agreement in order to limit emissions of greenhouse gases [116], they are not practical to be implemented for real-world road networks due to several reasons. The schemes require road users to go through a bidding or trading process before being able to use the paths, while in reality, road users often needs an instantaneous route choice decision. In addition, the schemes raise concerns about inequity when the poor is deprived of accessing to the infrastructure, and the rich holds most of the permits. Furthermore, although the tradable permit scheme approach, in some aspects, can reduce the computational burden of finding an allocation rule for very simple networks such as single OD pair networks, it is still not clear how to determine the allocation rule for heterogeneous travelers in general networks.

2.4.2 Algorithmic Mechanism Design

Since we are working on large-scale transportation networks, besides the game-theoretic aspects, it is also important to address computational tractability and scalability. The study of computational aspects in mechanism design has been investigated recently in an emerging sub-discipline called *algorithmic mechanism design*. On the one hand, researchers in this community address questions on how to bring computational aspects into economics and game theory by asking what equilibria notions are reasonable to assume. On the other hand, they also focus on how to bring game theory and economics into computer science and algorithmic theory by designing reasonable algorithms that are resilient to selfish behavior of agents [21].

Computer scientists have used the algorithmic mechanism design approach to tackle important problems in communication networks including resource allocation, pricing and multicast cost sharing [24–31]. As the Internet is rapidly growing in scale, the interaction between administrative domains known as Autonomous Systems and end-users have become extremely complicated due to different goals and incentives. In addition, since the network is shared by a large number of users, congestion may occur when resource demands exceed the capacity (e.g. link bandwidths), leading to packet delay and loss. Auctions have been suggested multiple times as an efficient mechanism that incentivizes users to share network resources in an optimal manner. For example, Lazar et al. devised a mechanism called Progressive Second Price (PSP) auction for allocating variable-size shares of a resource among a fix set of users [24]. The mechanism was proven to be incentive compatible and efficient in the sense that it maximizes the social welfare. Later, Maille et al. extended the PSP auction to deal with stochastic environment where users randomly enter and leave the system according to Poisson processes [25]. Other works on network resource allocation based on auction theory can be found in [26–28]. Although these works consider different possible mechanisms to resolve the incentive issues in Internet routing, the resulting routing protocols are hard to be implemented using the current network stack. In particular, the interdomain routing on the Internet is handled by the Border

Gateway Protocol (BGP), and it would be desirable to have a mechanism that can utilize this protocol. In [29], the authors propose a mechanism with a distributed allocation algorithm that is a straightforward extension to BGP and causes only modest increases in routing table size and convergence time. This mechanism aims to maximize network efficiency by routing packets along the lowest-cost paths.

2.4.3 Vickrey-Clarke-Groves Mechanisms

To achieve the incentive compatibility property, all of the above mechanisms for communication networks are constructed based on the arguably most important positive result in mechanism design, i.e., the generalized Vickrey-Clarke-Groves (VCG) mechanism [21, 117–119]. The main ideas of the VCG mechanism can be explained in the following scenario.

We assume that there are n players and a set of alternatives A that the mechanism can perform to allocate resources to all players. Each player i has his true valuation $v_i(a)$ for each alternative $a \in A$. In addition, each player has a reported valuation $b_i(a)$ for each alternative $a \in A$ that he reveals to the mechanism. If alternative a^* is executed, the utility of player i is equal to the difference between his true valuation for a^* and the cost that he needs to pay to the mechanism as follows:

$$u_i = v_i(a^*) - p_i(a^*).$$

The goal of the mechanism is to determine the alternative a^* that maximize social welfare $\sum_i v_i(a)$. First, based on the reported value $b_i(a)$ of each player i for each alternative a , the VCG mechanism find the alternative a^* that maximizes $\sum_i b_i(a)$. Then, the VCG mechanism charges each player i with price

$$p_i(a^*) = \left[\max_{a \in A} \sum_{j \neq i} b_j(a) \right] - \sum_{j \neq i} b_j(a^*). \quad (2.9)$$

The first component is obtained by removing player i and recomputing the maximum reported valuation of all other players that can be achieved with another alterna-

tive. The second component is the sum of reported valuation of all players except i when alternative a^* is chosen. Intuitively, the VCG payment rule implies that each player i needs to pay the externality that he imposes on others when joining the mechanism. We emphasize that this result coincides with the marginal cost pricing principle discussed in Section 2.3 under more general settings.

The VCG mechanism has several important properties as follows:

1. **Incentive compatibility (IC) property:** The mechanism is truthful in the sense that the best strategy for each player is to reveal his true valuation for each alternative. Let $u_i(v_i, b_{-i})$ denote the utility of player i when he reports his true valuation v_i , and $u_i(b_i, b_{-i})$ denote his utility when he reports b_i , given that all other players report b_{-i} . Then the IC property shows that $u_i(v_i, b_{-i}) \geq u_i(b_i, b_{-i})$ for all b_i . To see this property, we denote:

$$a^* = \arg \max_{a \in A} \left[b_i(a) + \sum_{j \neq i} b_j(a) \right] \text{ and } \hat{a}^* = \arg \max_{a \in A} \left[v_i(a) + \sum_{j \neq i} b_j(a) \right]$$

as the alternatives selected by the VCG mechanism when player i reports b_i and v_i respectively. We then have:

$$\begin{aligned} u_i(v_i, b_{-i}) &= v_i(\hat{a}^*) - p_i(\hat{a}^*) \\ &= v_i(\hat{a}^*) - \left[\left[\max_{a \in A} \sum_{j \neq i} b_j(a) \right] - \sum_{j \neq i} b_j(\hat{a}^*) \right] \\ &= \left[v_i(\hat{a}^*) + \sum_{j \neq i} b_j(\hat{a}^*) \right] - \left[\max_{a \in A} \sum_{j \neq i} b_j(a) \right] \\ &= \left[\max_{a \in A} \left[v_i(a) + \sum_{j \neq i} b_j(a) \right] \right] - \left[\max_{a \in A} \sum_{j \neq i} b_j(a) \right] \\ &\geq \left[v_i(a^*) + \sum_{j \neq i} b_j(a^*) \right] - \left[\max_{a \in A} \sum_{j \neq i} b_j(a) \right] \\ &= v_i(a^*) - \left[\left[\max_{a \in A} \sum_{j \neq i} b_j(a) \right] - \sum_{j \neq i} b_j(a^*) \right] \\ &= v_i(a^*) - p_i(a^*) = u_i(b_i, b_{-i}). \end{aligned}$$

2. **No-positive transfers property:** Each player's payment p_i is always non-negative, which means that the mechanism never pays to the players. This property can be seen clearly in Eq. 2.9.

3. **Individual rationality (IR) property:** If the valuation v_i of each player i is non-negative, then player i 's utility u_i is also non-negative for all i . This means players do not lose anything if they participate in the mechanism. We can see this property as follows:

$$\begin{aligned}
u_i(v_i, b_{-i}) &= \left[\max_{a \in A} \left[v_i(a) + \sum_{j \neq i} b_j(a) \right] \right] - \left[\max_{a \in A} \sum_{j \neq i} b_j(a) \right] \\
&\geq \left[\max_{a \in A} \sum_{j \neq i} b_j(a) \right] - \left[\max_{a \in A} \sum_{j \neq i} b_j(a) \right] = 0.
\end{aligned}$$

In spite of all these useful properties, much work has shown that in complex mechanism design problems such as combinatorial auctions, the allocation rule in the VCG mechanism is NP-hard to compute [120–122]. The key technical difficulty is that VCG mechanism requires an optimal solution to ensure the IC property, and this makes the mechanism computationally intractable. Many computationally tractable approximation algorithms or heuristics have been suggested, but they usually leads to untruthful mechanisms [120]. Resolving computational issues in VCG mechanisms while keeping the IC property of the mechanisms is still an active research area [121].

While the literature on algorithmic mechanism design for computer network routing is rich with a variety of research directions, applying algorithmic mechanism design for traffic routing in urban transportation networks remains unexplored. Although both types of networks share some similar modeling aspects, they are substantially different from each other in their inherent characteristics. In particular, in computer networks, utility functions of users are unknown to the network. Even if the utility functions are known, there is no central authority that knows all the link capacities and network topology [40]. In addition, autonomous systems (AS) in the network can run any routing algorithm that benefits them the most. Therefore, we can hardly sim-

ulate the flows for the entire computer networks such as the Internet. In contrast, in transportation networks, passenger utility functions can be modeled quite accurately based on regression analysis [41]. In addition, we can monitor real-time traffic flows via sensor networks [42–45] or community based mobile navigation applications such as Waze [46]. As a result, current models and results in computer network routing cannot be used and applied in urban transportation network routing.

In this thesis, we aim to design a mechanism that works specifically for transportation networks. To do this, we need to overcome several computational issues caused by time-varying travel demands from heterogeneous road users with private preference information on general networks. In particular, designing an allocation rule and an associated payment rule with the IC property is challenging due to the combinatorial nature of the problem. We, however, observe that instead of allocating paths for a number of passengers at the same time, we can allocate a path for each new passenger *sequentially*. Correspondingly, we can introduce a new appropriate equilibrium notation called *sequential social surplus*. In addition, travelers' utility function in transportation networks can be reliably derived based on results in disaggregate travel demand modeling that have been studied over the past four decades (see Chapter 2 in [104] and references therein). By utilizing these observations, we can overcome computational issues to design a strategy-proof mechanism in a proper sense. In particular, allocated paths are computed to match passengers' preference and to maximize sequential social surplus, and tolls are computed to ensure incentive-compatibility. Once all passengers are incentivized to commit to allocated paths, the complexity of predicting future traffic flows and computing paths and tolls for incoming passengers will be reduced significantly. In the following discussion, we will review important shortest path algorithms that are the building blocks of computing the allocation rule and payment rule in our proposed mechanism.

2.5 Shortest Path Algorithms

2.5.1 Unconstrained Shortest Path Problems

Given a network represented by graph $G = (V, E)$ with nodes V and directed links (also called arcs) E and the weights of all directed links are non-negative, the traditional shortest path problem is to build a *shortest path tree* from a single source node r_0 (also called root or origin) to all other nodes in the network. More specifically, the shortest path tree is a spanning directed tree of G , rooted at r_0 , which, for each node $v \in V$, contains a shortest path from r_0 to v [123]. All shortest path algorithms have a labeling process of finding a *label* for each node that contains the *cost from root* and the *predecessor node* on path from root to that node. To do this, the algorithms need two data structures: one for finding links out of each node and one for keeping track of candidate nodes to add to the shortest path tree, which is known as a *candidate list*. There are two types of shortest path algorithms including *label setting* and *label correcting* algorithms as described below.

The first efficient shortest path algorithm was proposed by Dijkstra in 1959 [124]. Dijkstra's algorithm is a label setting algorithm since when a link is added to the shortest path tree, it is kept permanently in the tree. Dijkstra's algorithm first adds the source node to the candidate list with label indicating zero cost. As the algorithm progresses, it chooses a node with the lowest label from the candidate list to process. When a node is taken out of the candidate list to process, the shortest path from root to the processed node has been found, and other labels in the candidate list are updated if it is faster to reach these nodes via the processed node. Neighboring nodes of the processed node are also put into the candidate list if they have not been in the list before. This procedure continues until all shortest paths have been found. In Dijkstra's algorithm, each node is taken out of the candidate list to add to the shortest path tree only once, and the candidate list can be implemented as a minimum priority queue or a minimum heap.

Dijkstra's algorithm is fastest for dense networks in which the average number of

links out of each node (or out-degree of each node) is greater than 30. For more sparse network in which the average out-degree is less than 30, a series of algorithms known as label correcting algorithms have been developed to reduce the computing time. In label correcting algorithms, a node taken out of the candidate list can be revisited later if better paths are found. The improvement of each algorithm depends on how the candidate list is managed. In particular, in Bellman-Ford algorithm developed in 1958, a new discovered node is always put on the back of the candidate list, and the next node is taken from the front of the list [125–127]. In 1974, Page devised a faster algorithm called D’Esopo-Pape by placing a new node in front of the candidate list if it has been on the list before, otherwise putting it on the back of the list [128]. Bertsekas, in 1992, proposed an algorithm that puts new node on the front of the candidate list if its label is smaller than current front node, otherwise puts it on the back of the list [129]. In the same year, Hao and Kocur suggested a faster algorithm that puts a new node on the front of the list if it has been on the list before. Otherwise, the new node is put on the back of the list if its label is greater than the front node label and on the front of the list if its label is smaller [130].

Theoretically, in the worst case scenario, label setting algorithms are faster than label correcting algorithms. Label setting algorithms run in $O(a^2)$ time in simple versions and in $O(a \lg n)$ time with a heap, where a is the number of links and n is the number of nodes. For label correcting algorithms, the worst-case time complexity is $O(2^a)$ (except that Bellman-Ford algorithm runs in $O(na)$ time [125]). In practice, however, label correcting algorithms, which run averagely in $O(a)$ time, usually outperform label setting algorithms, which run averagely in $O(a \lg n)$ time with a heap. This is because label correctors with an appropriate candidate list data structure in fact make very few corrections and run fast.

2.5.2 Constrained Shortest Path Problems

The basic problem of finding a shortest path between two specific nodes in a general network, as discussed above, can be solved in polynomial time by many efficient

labeling algorithms. Viewed from the mathematical programming perspective, the traditional shortest path problem can be modeled as an integer programming problem. Since this problem possesses the unimodularity property in this case, applying linear programming relaxation can solve the problem in polynomial time. In our mechanism, however, we not only want to compute a shortest path for each passenger, but we also want to make sure that the path matches with the passenger’s personal preference. To do this, we need to introduce an additional constraint to our shortest path algorithm. In fact, the constrained shortest path problems have been addressed by a number of researchers in the literature and can be classified into three main problems as discussed in [131]. The first problem is the *resource constrained shortest path* problem, which is the problem of a traveler with a budget of various resources who needs to find the quickest way to reach his destination without overspending his budget. The second is the *vertex constrained shortest path* problem, which is the problem of finding the shortest path that pass through a set of specified nodes (also called vertices). The last is the *time constrained shortest path problem* where the arc lengths are time-dependent and/or time windows exist for the nodes.

As our objective is to assign each passenger a path that matches with his travel time constraint, the induced shortest path problem in our mechanism is a time windows constrained shortest path problem. It has been showed that the shortest path problem with time windows constraints and possible negative edge costs is NP-hard [132, 133]. However, by leveraging the properties of transportation networks and the novel design of our mechanism, under some mild assumptions, we can apply the label-correcting algorithm to solve our fairness constrained shortest path (FCSP) problem in polynomial time. We will describe this process in detail in Chapter 4 and Chapter 5.

Chapter 3

Problem Formulation

This chapter develops a formal problem formulation for our game-theoretic approach to the optimal traffic routing problem. Section 3.1 provides preliminary notations and definitions used in the rest of the thesis. Following this, we present our assumptions in Section 3.2. Section 3.3 formulates a model for a general network with dynamic traffic flows of heterogeneous travelers. We present our User-Centric Dynamic Pricing (UCDP) Mechanism in detail in Section 3.4.

3.1 Preliminary Notations and Definitions

3.1.1 Graph

A graph $G = (V, E)$ consists of a finite set V of nodes (or vertices) and a finite set E of links (or edges, or arcs, or branches), which connect pairs of distinct nodes [134]. A link that connects node $x \in V$ and $y \in V$ is denoted as (x, y) . An origin-destination (OD) pair is denoted as $OD(r, s)$. When it is clear in a context, we use (r, s) instead of $OD(r, s)$ to refer to an OD pair.

A graph is called directed (or oriented) graph if every link in the graph has a specified orientation (or direction). In contrast, if no link has a specific orientation, the graph is undirected (or nonoriented). A directed link (x, y) has a direction from x to y and is usually drawn with an arrowhead that indicates its direction. Examples

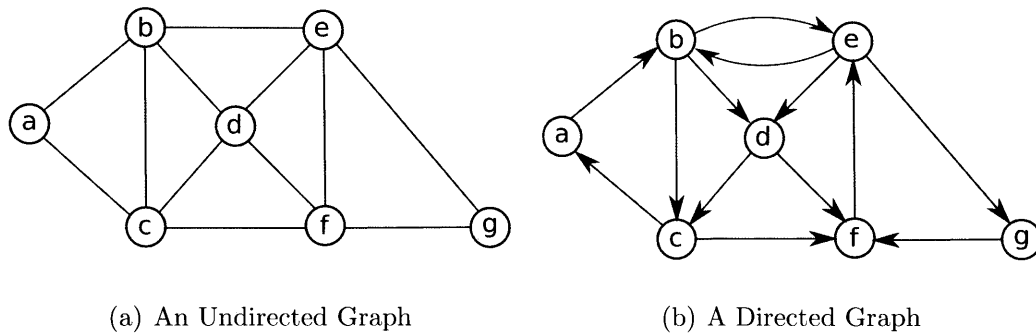


Figure 3-1: Examples of graphs.

of undirected and directed graphs are illustrated in Fig. 3.1.1.

A link is incident on the two nodes that it connects. Any two nodes connected by a link are called adjacent nodes. Similarly, two links are adjacent if they are connected by a node. For undirected graphs, the degree of a node is the number of links incident on it. For directed graphs, the indegree of a node is the number of links going to that node. The outdegree of a node in this case is the number of links going out from that node.

A path in a graph consists of a sequence of adjacent links that connect a sequence of adjacent nodes. A path begins at a node and ends at another node. In a directed graph, paths are also directed. A path can be represented as sequences of adjacent nodes such as $P = \{a, b, c, \dots, i, j, k\}$ or of adjacent links such as $P = \{(a, b), (b, c), \dots, (i, a), (j, k)\}$. If the starting node and ending node of a path coincide, the path is called a cycle (or circuit). A path is simple if each link appears only once in the sequence of links, and is elementary if each node is visited only once. In this thesis, we only consider simple and elementary paths.

3.1.2 Poisson Process

Suppose that there is a sequence of demand arrivals (or events) such that: (i) successive demand inter-arrival (or inter-event) times are mutually independent, and (ii) demand inter-arrival times are all described by the same exponential probability density function (PDF), then the number of arrivals constitutes a Poisson process.

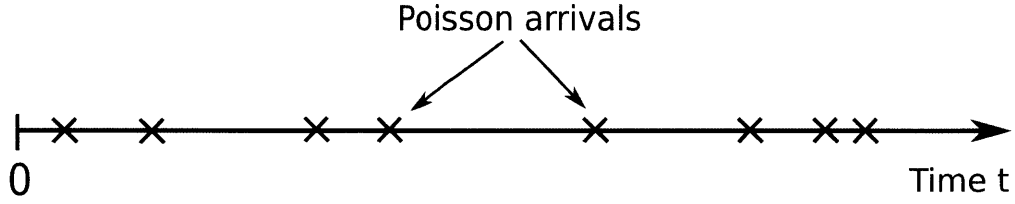


Figure 3-2: Illustration of Poisson arrivals in time.

Fig. 3.1.2 illustrates Poisson arrivals in time. The exponential PDF that represents demand inter-arrival times of a Poisson process is:

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

where λ is arrival rate measured by the average number of arrivals per unit of time, and is also referred to as the intensity of the arrivals of Poisson events. Let $N(t)$ denote the above Poisson process. Then the probability that n events occurring in the time interval $[0, t]$ is:

$$P(N(t) = n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}.$$

The mean and variance of $N(t)$ are both λt .

The Poisson process plays a central role in transportation modeling, as it can be used to model approximately the occurrence of random events such as the arrival of uncoordinated demands at a transportation facility, or the passage of cars by an observation point on a highway with free-flowing traffic.

3.1.3 Mechanism

Our mechanism has the following components and notations:

- A set $H = \{H_n\}_{n=1}^N$ of all heterogeneous passengers where H_n denotes a class of homogeneous passengers who have the same value of time $\bar{\alpha}_n$. If passenger i belongs to class H_n , we know that his value of time is $\alpha_i = \bar{\alpha}_n$.
- A set Θ that represents all possible types (i.e., preferences) of all passengers.

Each passenger i is required to report his type to the mechanism. Let $\theta_i \in \Theta$ denote passenger i 's true type and $\theta'_i \in \Theta$ denote his reported type.

- An allocation rule $\pi : \Theta \times \mathcal{X} \rightarrow \mathcal{P}$, where \mathcal{P} is the set of all paths that can be allocated to the set of passengers H , and \mathcal{X} is the set of all network states. In particular, a state $x = (k, q) \in \mathcal{X}$ is the state of the network at any time instant, where k is density, measured by number of vehicles per kilometer and q is the number of travelers on all links E of the network. The allocation rule computes a path p for a new passenger i based on his reported type θ'_i and the current network state $x \in \mathcal{X}$.
- A payment rule $w : \mathcal{P} \times \mathcal{X} \rightarrow \mathbb{R}$. The payment rule determines the toll that passenger i must pay to use the recommended path $p \in \mathcal{P}$.

3.2 Assumptions

We use the below assumptions in our UCDP mechanism.

1. Each vehicle has only one passenger.
2. Passengers are rational and risk-neutral, i.e., they always choose paths that maximize their utilities.
3. Passengers naturally want to join the mechanism. This assumption is called individual rationality (IR) in game theory terminology. We will discuss IR in Section 7.2.
4. Each passenger i 's value of time α_i can be derived from several demographic data such as tax and income.
5. If there exists a path $p \in \mathcal{P}$ between an OD pair (r, s) , then we assume that there also exists public transportation such as train or light rail between nodes r and s with infinite capacity.

6. Travel times on all roads in the network are positive and uniformly bounded away from zero.

3.3 Dynamic Network Model

We consider a general traffic network represented by a graph $G = (V, E)$ with nodes V and directed links E . We assume that there are L demand traffic flows in the network, denoted by $\{d_l\}_{l=1}^L$. Each demand flow d_l can be represented as a Poisson process with rate $\lambda_l > 0$ arriving at nodes R , where $R \subseteq V$. Each passenger from these L demand flows has a request to travel from an origin $r \in V$ to a destination $s \in V$. Between each OD pair (r, s) , there exists path $p \in P^{rs}$, where P^{rs} is the set of all feasible paths between r and s . We define \mathcal{P} as the set of all P^{rs} .

We recall that $x = (k, q) \in \mathcal{X}$ represents the state of the network at any time instant. In particular, we have $k = (k_e)_{e \in E}$ and $q = (q_e)_{e \in E}$ in which k_e and q_e denote the density and the number of travelers on a particular link $e \in E$ at that time instant. Each link $e \in E$ has a jam density K_{J_e} at which the traffic on the link is congested with zero traffic flow. Based on the Greenshields model described at the beginning of Chapter 2, $K_{C_e} = \frac{1}{2}K_{J_e}$ denotes a critical density at which additional input of vehicles on the link decreases traffic flow and eventually leads to traffic congestion. The set of all feasible density on link $e \in E$ is K_e . We define $K = \prod_{e \in E} K_e$ so that $k \in K$. Similarly, the set of all feasible number of people on link e is Q_e , and $q \in Q = \prod_{e \in E} Q_e$. We note that x is time-varying.

At any time instant t , assuming that the network state is x , we would like to simulate the network state $\hat{x} \in \mathcal{X}$ after a path $p \in \mathcal{P}$ is assigned to a new traveler given that all current travelers in the network are following their assigned paths. Ideally, we would like to predict \hat{x} dynamically at any time instant in the future after the path p is assigned to the new passenger. However, to simplify the notations in the following discussion, we assume that from that time instant t , the state x of the network will not change until the new passenger completes his trip. We can simulate \hat{x} by simply adding the new traveler on each link $e \in p \subset E$ as if he were traveling on

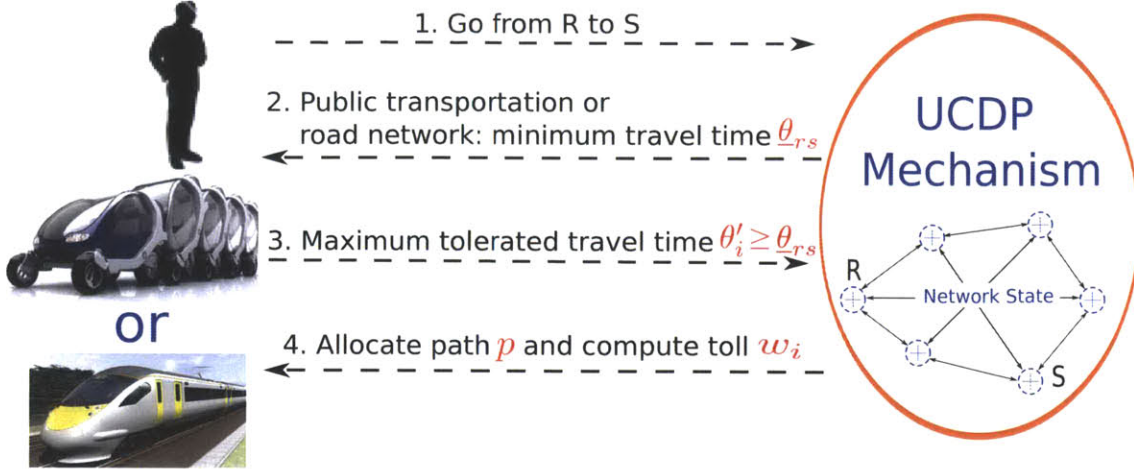


Figure 3-3: User-Centric Dynamic Pricing mechanism.

e at time t . Thereafter, given density on link e , the space-average speed of passengers traveling on the link can be computed by using the Greenshields model described in Chapter 2. In particular, $\hat{x} = \mathcal{G}(x, p)$ where $\mathcal{G} : \mathcal{X} \times \mathcal{P} \rightarrow \mathcal{X}$. We will return to dynamic and real-time prediction in Chapter 5.

The travel time on each link e , denoted by $\tau_e(\hat{k}_e)$, is a strictly increasing and convex function of density \hat{k}_e , where \hat{k}_e is extracted from $\hat{x} = (\hat{k}, \hat{q}) = \mathcal{G}(x, p)$ and is the experienced density of the driver. For a new driver who follows a path $p \in \mathcal{P}$, his experienced travel time along this path would be:

$$\tau(x, p) = \sum_{e \in p} \tau_e(\hat{k}_e). \quad (3.1)$$

3.4 User-Centric Dynamic Pricing Mechanism

We now discuss the user-centric dynamic pricing (UCDP) mechanism in detail. We remind that $H = \{H_n\}_{n=1}^N$ is the set of all heterogeneous passengers where H_n denotes a class of homogeneous passengers who have the same value of time $\bar{\alpha}_n$. If passenger i belongs to class H_n , we know that his value of time is $\alpha_i = \bar{\alpha}_n$. The UCDP mechanism is illustrated in Fig. 3-3 and works as follows. At time t , a passenger $i \in H_n$ who is the latest incoming passenger from L demand flows wants to travel from $r \in V$

to $s \in V$. Passenger i will report his maximum accepted travel time $\theta_i \geq \underline{\theta}_{rs}$, to a central system. In game-theory terminology, we define θ_i as the type of passenger i . The lower bound $\underline{\theta}_{rs}$ is the minimum time to travel from r to s and is announced publicly by the road authority. The value of $\underline{\theta}_{rs}$ is computed in Chapter 5. If there is currently no path to travel from r to s due to congestion, we define $\underline{\theta}_{rs} = +\infty$, and the passenger is advised to use public transportation. Otherwise, our mechanism then computes a suitable path p for this passenger and a toll $w_i(x, p)$ that he has to pay to complete his trip. The utility function of passenger i for such an assignment is determined as follows:

$$u_i(x, p, \theta_i) = \alpha_i(\theta_i - \tau(x, p)) - w_i(x, p), \quad (3.2)$$

where $\tau(x, p)$ is the travel time that passenger i spends when traveling along path p from r to s and can be computed by Eq. 3.1. We assume that passenger i is risk-neutral and seeks to maximize his utility $u_i(x, p, \theta_i)$ of the trip. Clearly, the first term in Eq. 3.2 can be considered as passenger i 's benefit of making the trip, and the second term is his cost. Thus, we denote $v_i(x, p, \theta_i) = \alpha_i(\theta_i - \tau(x, p))$ as passenger i 's benefit.

The road authority, on the other hand, tries to maximize the social surplus of all current travelers in the road network. When a path p is assigned to passenger i , the road authority's utility function is defined as follows:

$$V(x, p) = - \sum_{e \in E} \alpha_0^{i,e} \tau_e(\widehat{k}_e) q_e, \quad (3.3)$$

where $\alpha_0^{i,e}$ is the authority's value of time for passenger i on link e , which depends on this passenger's class, and \widehat{k}_e , extracted from $\widehat{x} = \mathcal{G}(x, p)$, is the new density on link e when passenger i travels along path p . In particular, if $e \notin p$, we have $\widehat{k}_e = k_e$, and if $e \in p$, \widehat{k}_e is the induced density due to $q_e + 1$ travelers on link e . Equation 3.3 implies that the road authority is better off if the total travel time of current travelers is smaller.

As we are handling dynamic traffic network in which demands arrive as Poisson processes, the ordinary social optimal concept is not suitable. Thus, we propose *sequential social surplus* as an alternative measurement of total society's benefit. More precisely, we define sequential social surplus as follows:

Definition 1 (Sequential Social Surplus) *Sequential social surplus is the sequence of the sum of road authority utility $V(x, p)$ and a new traveler's benefit $v_i(x, p, \theta_i)$ over time.*

Our goal is to assign each passenger i a path p that satisfies his travel time constraint θ'_i while maximizing sequential social surplus when passenger i joins the network. In addition, we want to ensure that the density on each link $e \in E$ does not exceed its critical density K_{C_e} for all times t , i.e., the network is always congestion-free. The exceeding demand is rerouted to public transportation. While we assume that each passenger will follow the assigned path, there is a concern that a passenger i could increase his utility $u_i(x, p, \theta_i)$ if he lies about his true tolerated travel time θ_i . Such manipulation would negatively affect the overall network efficiency. To avoid this situation, our approach, described in Chapters 4 and 5, computes a path allocation rule and a payment rule that are incentive compatible (or strategy-proof), which means passengers have no incentive to lie about their preference.

Chapter 4

Approach

This chapter presents our approach to design an allocation rule and a payment rule, which are two important components in our mechanism. Section 4.1 describes the allocation rule, which assigns a path that satisfies each user's time preference while maximizing sequential social surplus and ensuring that the network is not congested at any time instant. The allocation rule is formulated as a fairness constrained shortest path (FCSP) problem. The payment rule, which is described in Section 4.2, calculates a toll that each passenger pays to use his assigned path based on the marginal cost principle. Section 4.2 also presents important theorems, in which we prove that the mechanism has incentive compatibility and no-positive transfer properties under the above allocation rule and payment rule.

4.1 Allocation Rule

We recall that our first objective is to achieve a path allocation rule that matches with each user's preference while maximizing sequential social surplus and avoiding congestion. More precisely, given that at time t , a passenger i reports his type θ'_i , we aim to solve the following optimization problem for all θ'_i :

$$\max_{p \in P^{rs}} V(x, p) + v_i(x, p, \theta'_i)$$

subject to

$$\tau(x, p) \leq \theta'_i, \quad (4.1)$$

$$\widehat{k}_e \leq K_{c_e}, \quad \forall e \in E. \quad (4.2)$$

The objective function is to maximize sequential social surplus, which is the sum of road authority utility $V(x, p)$ and passenger i 's benefit $v_i(x, p, \theta_i)$. The preference constraint in Eq. 4.1 expresses an important requirement that the experienced travel time of passenger i along his assigned path does not exceed his reported maximum tolerated travel time. To prevent congestion, the link density constraint in Eq. 4.2 dictates that the density of link e when passenger i reaches that link must be lower than its critical density.

The objective function can be rewritten as follows:

$$\begin{aligned} & V(x, p) + v_i(x, p, \theta'_i) \\ &= - \sum_{e \in E} \alpha_0^{i,e} \tau_e(\widehat{k}_e) q_e + \alpha_i (\theta'_i - \tau(x, p)) \\ &= - \left[\sum_{e \in E \setminus p} \alpha_0^{i,e} \tau_e(k_e) + \sum_{e \in p} \alpha_0^{i,e} \tau_e(\widehat{k}_e) \right] q_e + \alpha_i \theta'_i - \alpha_i \tau(x, p) \\ &= - \sum_{e \in E \setminus p} \alpha_0^{i,e} \tau_e(k_e) q_e + \alpha_i \theta'_i - \sum_{e \in p} \alpha_0^{i,e} \tau_e(\widehat{k}_e) q_e - \alpha_i \sum_{e \in p} \tau_e(\widehat{k}_e) \\ &= - \sum_{e \in E \setminus p} \alpha_0^{i,e} \tau_e(k_e) q_e + \alpha_i \theta'_i - \sum_{e \in p} \tau_e(\widehat{k}_e) (\alpha_0^{i,e} q_e + \alpha_i) \\ &= - \sum_{e \in E} \alpha_0^{i,e} \tau_e(k_e) q_e + \alpha_i \theta'_i - \sum_{e \in p} \left\{ \left[\tau_e(\widehat{k}_e) - \tau_e(k_e) \right] \alpha_0^{i,e} q_e + \tau_e(\widehat{k}_e) \alpha_i \right\}, \end{aligned}$$

where q_e is the number of travelers in each link right before passenger i starts his trip.

Since we know the network state x and passenger i 's reported type as well as his value of time, the terms $\sum_{e \in E} \alpha_0^{i,e} \tau_e(k_e) q_e$ and $\alpha_i \theta'_i$ are constant. Therefore, our problem becomes a fairness constrained shortest path (FCSP) with bounded travel

time:

$$\min_{p \in Prs} \sum_{e \in p} \left\{ \left[\tau_e(\widehat{k}_e) - \tau_e(k_e) \right] \alpha_0^{i,e} q_e + \tau_e(\widehat{k}_e) \alpha_i \right\} \quad (4.3)$$

subject to

$$\tau(x, p) \leq \theta'_i, \quad (4.4)$$

$$\widehat{k}_e \leq K_{Ce}, \quad \forall e \in E. \quad (4.5)$$

The equivalent arc-based formulation is as follows:

$$\min_z \sum_{e \in E} z_e \left\{ \left[\tau_e(\widehat{k}_e) - \tau_e(k_e) \right] \alpha_0^{i,e} q_e + \tau_e(\widehat{k}_e) \alpha_i \right\} \quad (4.6)$$

Subject to

$$\sum_{e \in r^+} z_e = 1, \quad \sum_{e \in s^-} -z_e = -1, \quad (4.7)$$

$$\sum_{e \in v^+} z_e - \sum_{e \in v^-} z_e = 0, \quad \forall v \in V \setminus \{r, s\}, \quad (4.8)$$

$$\sum_{e \in E} z_e \left[\tau_e(\widehat{k}_e) \right] \leq \theta'_i, \quad (4.9)$$

$$\widehat{k}_e \leq K_{Ce}, \quad \forall e \in E, \quad (4.10)$$

$$z_e \in \{0, 1\}, \quad \forall e \in E. \quad (4.11)$$

The objective function in Eq. 4.6 is to minimize the sum of the increment in total generalized travel time of current passengers on the roads and the generalized travel time of the new passenger. Here z_e is a binary variable, i.e., it is 1 if link e is on path p assigned for passenger i , and zero otherwise. The coefficients of z_e in the objective function are called edge costs. The first two constraints in Eq. 4.7 are the condition that passenger i needs to make a trip from r to s , where r^+ is the set of links going out from origin r and s^- is the set of links coming in destination s . Eq. 4.8

is a node balance constraint equation, where v^+ represents the set of links coming in node $v \in V \setminus \{r, s\}$ and v^- represents the set of links going out from v . The constraint in Eq. 4.9 ensures that the travel time of passenger i along his assigned path should not exceed his stated maximum accepted travel time. The coefficients of z_e in this constraint are called edge weights. The constraint in Eq. 4.10 restricts on link density to guarantee that congestion will not happen. We note that, to further utilize the roads, the upper bound K_{Ce} in Eq. 4.10 can be set to a higher value without changing the two algorithms presented in Chapter 5. If the preference constraint in Eq. 4.9 and the link density constraint in Eq. 4.10 are omitted, an optimal solution to this formulation would coincide with the social optimum in traditional traffic assignment at this time instant.

In addition, it is worth noting that our objective function also demonstrates how the value of time of passenger i can affect his assigned path. As the ratio of α_i to $\alpha_0^{i,e} q_e$ becomes extremely large, i.e., passenger i needs to reach the destination urgently, the road authority can deliberately neglect his value of time and can only consider to minimize travel time on each link for this passenger. Hence, the formulation has significant meaning in real-world routing when we need to provide a police car or an ambulance with the fastest path to reach the accident site. Therefore, the combination of the preference constraint and the role of passenger's value of time in the objective function successfully model fairness requirement for individuals. To the best of our knowledge, the proposed model is the first in the literature that considers all of the above aspects.

Regarding computational complexity of the problem, it is well known that the general shortest path problems with additional constraints and possible negative edge costs are NP-hard [133]. However, by leveraging the properties of transportation networks and the novel design of our mechanism, we can naturally handle the problem by assuming that travel times on all roads in the network are positive and uniformly bounded away from zero. Under these mild assumptions, our FCSP problem can be solved in polynomial time by using the label-correcting algorithm [132, 135]. The implementation of label-correcting algorithm will be explained in detail in Chapter 5.

4.2 Payment Rule

We now discuss how to design the payment rule such that the proposed mechanism satisfies incentive compatibility. Given that we have solved the FCSP problem in Eq. 4.3-4.5, with θ'_i , we obtain an optimal solution $p^*(\theta'_i)$, which is the path that we will allocate to passenger i . With this optimal path, we can calculate the payment that passenger i pays for using the path based on the marginal cost principle:

$$\begin{aligned} w_i(x, p^*(\theta'_i)) &= V(x, \emptyset) - V(x, p^*(\theta'_i)) \\ &= \sum_{e \in p^*(\theta'_i)} \left[\tau_e(\widehat{k}_e) - \tau_e(k_e) \right] \alpha_0^{i,e} q_e. \end{aligned} \quad (4.12)$$

The first term on the right side of Eq. 4.12 is the maximum value of the authority's utility when passenger i is not included, and the second term is the authority's utility when passenger i travels along path $p^*(\theta'_i)$. We note that tolls for different classes of passengers are different, depending on the authority's value of time $\alpha_0^{i,e}$ for a particular class. This payment rule would favor low-income groups and thus provides fairness among individuals as discussed by [15]. Under this price, the utility of passenger i , $u_i(x, p^*(\theta'_i), \theta_i)$, in Eq. 3.2 becomes:

$$\alpha_i(\theta_i - \tau(x, p^*(\theta'_i))) - V(x, \emptyset) + V(x, p^*(\theta'_i)). \quad (4.13)$$

The pricing mechanism we propose looks similar to the celebrated Vickrey-Clarke-Groves (VCG) mechanism that is incentive compatible (see Section 2.4.3). Nevertheless, our mechanism is considerably different from VCG mechanism because at each "game", we only consider a single player, i.e., a passenger, who indirectly plays with all current passengers in the network before actually joining the network. Moreover, we have additional preference and capacity constraints in our allocation rule problem, while VCG mechanism does not have. Therefore, it is necessary to prove that our mechanism still maintains the strategy-proof property in spite of the modifications as follows.

Theorem 1 (Incentive compatibility property) *Given the allocation rule determined by the FCSP problem and the marginal cost payment rule as in Eq. 4.12, reporting $\theta'_i = \theta_i$ is a weakly dominant strategy of passenger i , $\forall i \in H$.*

Proof: Assume that truth telling is not a weakly dominant strategy for some i , i.e. there exists some θ_i and θ'_i such that

$$u_i(x, p^*(\theta'_i), \theta_i) > u_i(x, p^*(\theta_i), \theta_i), \quad (4.14)$$

$$\tau(x, p^*(\theta'_i)) \leq \theta_i, \quad (4.15)$$

and the network is congestion-free under both reports θ_i and θ'_i . Equation 4.15 means that $p^*(\theta'_i)$ is a feasible solution to the optimization problem:

$$\max_{p \in P^{rs}} V(x, p) + v_i(x, p, \theta_i)$$

subject to

$$\begin{aligned} \tau(x, p) &\leq \theta_i, \\ \widehat{k}_e(e) &\leq K_{ce}, \quad \forall e \in E. \end{aligned}$$

We have $p^*(\theta_i)$ as the optimal solution to the above problem due to the allocation rule. Thus:

$$V(x, p^*(\theta'_i)) + v_i(x, p^*(\theta'_i), \theta_i) \leq V(x, p^*(\theta_i)) + v_i(x, p^*(\theta_i), \theta_i). \quad (4.16)$$

On the other hand, substituting Eq. 4.13 into Eq. 4.14, this yields

$$\begin{aligned} &\alpha_i(\theta_i - \tau(x, p^*(\theta'_i))) - V(x, \emptyset) + V(x, p^*(\theta'_i)) \\ &> \alpha_i(\theta_i - \tau(x, p^*(\theta_i))) - V(x, \emptyset) + V(x, p^*(\theta_i)) \end{aligned}$$

$$\begin{aligned} &\Leftrightarrow V(x, p^*(\theta'_i)) + v_i(x, p^*(\theta'_i), \theta_i) \\ &> V(x, p^*(\theta_i)) + v_i(x, p^*(\theta_i), \theta_i). \end{aligned}$$

Comparing the above inequality with Eq. 4.16, we reach a contradiction.

Similar to the VCG mechanism, our UC DP mechanism also has the no-positive transfers property as shown in the following theorem.

Theorem 2 (No-positive transfer property) *Given the allocation rule determined by the FCSP problem and the marginal cost payment rule as in Eq. 4.12, passenger i 's payment w_i is always non-negative, $\forall i \in H$. This means the mechanism never pays a positive payment to passengers.*

Proof: By definition in Eq. 4.12, the payment that passenger i pays for using a path assigned to him is

$$w_i(x, p^*(\theta'_i)) = \sum_{e \in p^*(\theta'_i)} \left[\tau_e(\widehat{k}_e) - \tau_e(k_e) \right] \alpha_0^{i,e} q_e.$$

According to the density-speed relationship in the Greenshields model described in Section 2.1, we can derive the travel time on link e when passenger i join that link $\tau_e(\widehat{k}_e)$ as follows:

$$\tau_e(\widehat{k}_e) = \frac{\mathcal{L}_e}{v_e^{max} \left(1 - \frac{\widehat{k}_e}{K_{J_e}} \right)},$$

where \mathcal{L}_e is the physical length of link e . A more detailed description of this equation is included in Section 5.1 in the next chapter. Given that the link length \mathcal{L}_e , the maximum speed v_e^{max} and the jam density K_{J_e} are constant, it is clear that the travel time $\tau_e(\widehat{k}_e)$ is a monotonically increasing function with respect to the link density \widehat{k}_e . As $\widehat{k}_e > k_e$, $\tau_e(\widehat{k}_e) > \tau_e(k_e)$. In addition, the road authority's value of time $\alpha_0^{i,e}$ and the flow on link e q_e are always non-negative. Therefore, $w_i(x, p^*(\theta'_i)) \geq 0$.

Chapter 5

UCDP Computation

In this chapter, we describe in detail how we allocate paths and compute tolls dynamically. Section 5.1 shows how network states are simulated by applying the Greenshields model. Section 5.2 provides an algorithm to compute minimum travel times announced by a road authority by using Dijkstra's algorithm. Following this, an algorithm for computing dynamic path allocation and dynamic tolls are presented in Section 5.3.

5.1 Traffic Simulation

Assuming that travelers in the networks are following assigned paths, we use the Greenshields model presented in Section 2.1 to simulate a network state at any time instant. The density k_e on each link e can be computed by counting the number of passengers on that link. Given maximum speed v_e^{max} and jam density K_{Je} on each link $e \in E$, the space-average speed v_e and travel time $\tau_e(\hat{k}_e)$ on each link e at each time instant are computed as:

$$v_e = v_e^{max} \left(1 - \frac{\hat{k}_e}{K_{Je}}\right),$$

and

$$\tau_e(\hat{k}_e) = \frac{\mathcal{L}_e}{v_e},$$

where \mathcal{L}_e is the physical length of link e . The traffic flow and the maximum traffic flow on each link e at each time instant are

$$f_e = v_e \hat{k}_e,$$

and

$$F_e^{max} = \frac{1}{4} v_e^{max} K_{Je},$$

respectively.

The above computation process can also be used to predict future network states as follows. From the current network state x , as passengers in the network are following their assigned paths, we can simulate to obtain their positions after some duration t . Hence, we can estimate the number of passengers on all links in the future. We now can apply the Greenshields model as above to compute the new network state.

Therefore, we assume that there exists a procedure `Greenshields`(x, t) that predicts the future network state from the current network x after a period of time t . We note that the more accurate prediction of $\hat{x} = \mathcal{G}(x, p)$ in Chapters 3-4 can be done by calling the procedure `Greenshields` before accessing an edge state (see Algorithm 1 and Algorithm 2). The computed states are real-time and dynamic.

5.2 Minimum Travel Time Computation

We compute minimum travel times using Dijkstra's algorithm on the graph with time-varying edge costs as shown in Algorithm 1. The algorithm takes the current network state x and the OD pair (r, s) as input and returns the minimum time to travel from r to s .

The algorithm is implemented using a minimum priority queue to store candidate vertices. Each vertex has a label (T, pre) where T is the time to reach that vertex from the origin r , and pre denotes the previous vertex on the shortest path from r to that vertex. When a vertex is taken out of the candidate list, we call the procedure `Greenshields` (Line 11) to update a new network state \hat{x} after taking the

shortest path from the origin r to that vertex. The predicted edge cost $\hat{\tau}_e$ (Line 12) for link $e \in E$ is $\tau_e(\hat{k}_e)$ where \hat{k}_e is computed at the time the passenger reach link e from the new network state \hat{x} . This computation is performed in the procedure `ComputeEdgeTravelTimes`. In addition, in this algorithm, a link with density greater than the upper bound in Eq. 4.10 is considered to be unavailable (see Line 14). We update vertex labels in Lines 15-18 if it is better to reach a vertex via the currently processed candidate vertex. We assume that simulation for obtaining predicted network states can be performed in constant time. Under this assumption, the running time of Algorithm 1 is $|E| \ln |V|$.

5.3 Dynamic Allocation and Pricing Algorithm

Algorithm 2 shows how assigned paths and tolls are computed dynamically from an origin node $r \in V$ to a destination node $s \in V$. The algorithm takes a reported maximum travel time θ' from a passenger as well as his value of time α as inputs. The algorithm also takes the road authority's value of time α_0 for this passenger on all links as an input. We use a label-correcting algorithm to find all Pareto-optimal paths. The details of the basic version of this label-correcting algorithm can be found in [132, 135].

Essentially, at each vertex, we maintain a set of labels (C, T, pre) , each of them consisting of a cost component C , a travel time component T , and a pointer pre . In particular, at vertex $v \in V$, the cost component of a label is the cost-of-arriving from r to v as defined in the FCSP problem (see Chapter 4) by following the path induced by the label. Similarly, the travel time component is the total travel time from r to v . The pointer component pre has the form (i, k) where i the previous vertex in the induced path, and k is the label index of the previous vertex i . At a vertex i , $b(i)$ denotes the number of labels has been constructed and maintained by the algorithm.

The algorithm uses a minimum priority queue Q (Line 2), which is sorted by the cost component, to store all labels. In the main loop of the algorithm, we first extract a label (C_i^m, T_i^m, pre_i^m) with the minimum cost component to process (Line 7). Here,

we have i is the processed vertex, and m is the index of the label at this vertex. Similar to Algorithm 1, we estimate the future network state \hat{x} (see Line 8) if the passenger travels the path induced by the processed label to reach the processed vertex. At Line 9, we compute the edge costs, denoted as \hat{c}_e , and the edge travel times, denoted as $\hat{\tau}_e$, from the FCSP problem in the procedure `ComputeEdgeCostTravelTimeFCSP` as follows:

$$\hat{\tau}_e = \tau_e(\hat{k}_e), \text{ and } \hat{c}_e = \left[\hat{\tau}_e - \tau_e(k_e) \right] \alpha_0^e q_e + \hat{\tau}_e \alpha.$$

We then consider all neighbor vertices j of the processed vertex i such that edge (i, j) is not congested, and time to reach vertex j does not exceed θ' from Line 11 to Line 23. In Lines 13-15, we check if there exists a label at vertex j that outperforms the path to reach j via i . Such label (C_j^n, T_j^n, pre_j^n) exists if

$$C_i^m + \hat{c}_{i,j} \geq C_j^n \quad \text{and} \quad T_i^m + \hat{\tau}_{i,j} \geq T_j^n.$$

When no such label exists, we construct a new label for the vertex j in Lines 17-18. If j is not the destination vertex s , we put the new label into the priority queue Q (see Lines 19-20). To expediting the running time of the algorithm, we remove from the queue those labels at vertex j that are dominated by the new label in Lines 21-23.

Finally, the path with smallest cost from r to s and the associated toll is computed at Lines 24-25. When algorithm finishes processing the entire queue Q , at the destination vertex s , we obtain all Pareto-optimal paths from r to s that satisfy the passenger's preference. We can return a list of paths and corresponding tolls for users to choose from. Nevertheless, based on our model, a rational user would choose the path with smallest cost C_s as currently returned in Algorithm 2 since this path maximizes the user's utility. The procedure `GetBestPath` returns the best path p^* by tracing previous pointers backwards from s to r . With the return path p^* , the procedure `MarginalPricing` computes the associated toll as in Eq. 4.12. The complexity of this algorithm is $O(\theta'^2|V|^2)$ where θ' is the reported maximum tolerated time and $|V|$ is the number of nodes [132, 135].

Algorithm 1: Dijkstra($G = (V, E), x, r, s$)

```
// Initialization
1  $(T_r, pre_r) \leftarrow (0, \emptyset)$  ;
   //  $Q$ : minimum priority queue sorted by  $T_i$ 
2  $Q \leftarrow \{(T_r, pre_r)\}$  ;
3  $InQ_r \leftarrow 1$  ;
4 for  $i \in V \setminus \{r\}$  do
5    $(T_i, pre_i) \leftarrow (\infty, \emptyset)$  ;
6    $InQ_i \leftarrow 0$  ;

// Computing shortest path
7 while  $Q \neq \emptyset$  do
8    $(T_i, pre_i) \leftarrow \text{Pop}(Q)$  ;
   // If destination  $s$  is found
9   if  $i == s$  then
10     $\text{break}$ ;

   //  $T_i$ : shortest time to reach  $i$  from  $r$ 
   // Estimate network state after reaching  $i$ 
11  $\hat{x} \leftarrow \text{Greenshields}(x, T_i)$ ;

   // Compute time-varying edge travel times
12  $\hat{\tau} \leftarrow \text{ComputeEdgeTravelTimes}(\hat{x})$ ;

   // Update labels
13 for  $j \in \text{Neighbor}(G, i)$  do
14   if  $\hat{k}_{i,j} \leq K_{C(i,j)} \wedge T_i + \hat{\tau}_{i,j} < T_j$  then
15      $(T_j, pre_j) \leftarrow (T_i + \hat{\tau}_{i,j}, i)$  ;
16     if  $InQ_j == 0$  then
17        $Q \leftarrow Q \cup \{(T_j, pre_j)\}$ ;
18        $InQ_j \leftarrow 1$  ;

19 return  $T_s$ 
```

Algorithm 2: UC DP($G = (V, E), x, r, s, \theta', \alpha, \alpha_0$)

```
// Initialization
1  $(C_r^1, T_r^1, pre_r^1) \leftarrow (0, 0, \emptyset)$ ;
   //  $Q$ : minimum priority queue sorted by  $C_i^m$ 
2  $Q \leftarrow \{(C_r^1, T_r^1, pre_r^1)\}$ ;
   //  $b(i)$ : number of constructed labels at vertex  $i$ 
3 for  $i \in V \setminus \{r\}$  do
4    $b(i) \leftarrow 0$ ;
5  $b(r) \leftarrow 1$ ;

   // Computing assigned path
6 while  $Q \neq \emptyset$  do
7    $(C_i^m, T_i^m, pre_i^m) \leftarrow \text{Pop}(Q)$ ;
   //  $T_i^m$ : time to reach  $i$  from  $r$  with  $m^{\text{th}}$  path
   // Estimate network state after reaching  $i$  using this path
8    $\hat{x} \leftarrow \text{Greenshields}(x, T_m)$ ;
   // Compute time-varying edge costs and travel times for FCSP
9    $(\hat{c}, \hat{\tau}) \leftarrow \text{ComputeEdgeCostTravelTimeFCSP}(\hat{x})$ ;
10  for  $j \in \text{Neighbor}(G, i)$  do
11    if  $T_i^m + \hat{\tau}_{i,j} \leq \theta' \wedge \hat{k}_{i,j} \leq K_{C(i,j)}$  then
12       $flag \leftarrow 1$ ;
13      for  $n = 1 \rightarrow b(j)$  do
14        if  $C_i^m + \hat{c}_{i,j} \geq C_j^n \wedge T_i^m + \hat{\tau}_{i,j} \geq T_j^n$  then
15           $flag \leftarrow 0$ ;
16      if  $flag == 1$  then
17         $b(j) \leftarrow b(j) + 1$ ;
18         $(C_j^{b(j)}, T_j^{b(j)}, pre_j^{b(j)}) \leftarrow (C_i^m + \hat{c}_{i,j}, T_i^m + \hat{\tau}_{i,j}, (i, m))$ ;
19        if  $j \neq s$  then
20           $Q \leftarrow Q \cup \{(C_j^{b(j)}, T_j^{b(j)}, pre_j^{b(j)})\}$ ;
21        for  $(C_j^n, T_j^n, pre_j^n) \in Q$  do
22          if  $C_j^n \geq C_j^{b(j)} \wedge T_j^n \geq T_j^{b(j)}$  then
23             $Q \leftarrow Q \setminus \{(C_j^n, T_j^n, pre_j^n)\}$ ;

24  $p^* \leftarrow \text{GetBestPath}(C_s, A_s, pre_s)$ ;
   // Computing toll
25  $w \leftarrow \text{MarginalPricing}(x, p^*)$ ;
26 return  $(p^*, w)$ 
```


Chapter 6

Experiments

This chapter demonstrates the performance of the proposed mechanism through two simulated routing experiments. Section 6.1 presents the first experiment, in which we simulate the UCDP mechanism on a parallel-link network with one OD pair. We then compare the performance of the UCDP mechanism with user-equilibrium and social-optimal performance. In the second experiment described in Section 6.2, we consider a general network with multiple OD pairs including a bottle neck. We demonstrate how the road authority uses the UCDP mechanism to manage traffic flows when disruption occurs at the bottle neck.

6.1 Parallel-link Network

In the first experiment, we tested the UCDP mechanism on a parallel-link network with one OD pair, in which links (0,3) and (3,4) are longer than the other links, as shown in Fig. 6-1. The specifications for this network including physical length, maximum speed, critical density, and maximum flow for all links are shown in Table 6.1. The total maximum traffic flow (or capacity) on the three routes from node 0 to node 4 is 1187.5 vehicles per hour (v/h). Passengers can also use public transportation to reach the destination in one hour. The value of times for all passengers are 10 USD/h, and the value of time of the authority, $\alpha_0^{i,e}$, is 100 USD/h for all links. At

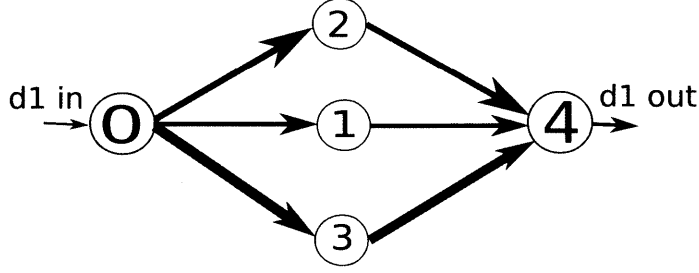


Figure 6-1: Parallel-link network.

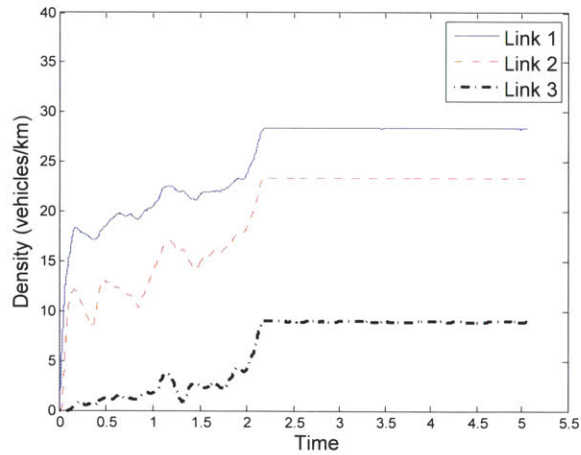
Table 6.1: Network specifications for Fig. 6-1. Capacity 1187.5 v/h

Link	\mathcal{L} (km)	v^{max} (km/h)	K_c (veh./km)	F^{max} (veh./h)
01, 14	2.5	60.0	15.0	450.0
02, 24	3.5	70.0	12.5	437.5
03, 34	10.0	120.0	5.0	300.0

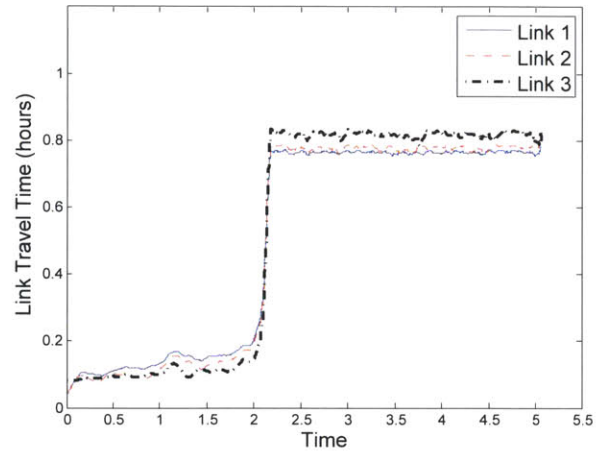
first, the flow rate λ_1 of the traffic demand d_1 is set at 970 vehicles per hour (v/h), which is lower than the total capacity of the network. We then increase the demand to 1500 v/h to observe the network performance when the demand exceeds the capacity. In these cases, we compare the performance of the UCDP mechanism with user-equilibrium (UE) and social-optimal (SO) performance. Since we are dealing with dynamic demand flows, we can not compute *a priori* UE and SO performance as usually done in static traffic assignment. Therefore, UE and SO performance in Figs. 6.1, 6-3 and 6-5 is computed *sequentially* for every new arriving passenger.

When the demand is 970 v/h, as shown in Fig. 6.1, under the UE setting with no toll, the densities, link travel times, and flows and passengers' travel time for all links quickly reach their critical values after about 2.2 hours. This is because all passengers act selfishly to choose paths with minimum travel time at the time they arrive at the origin. Therefore, the road reaches the congestion condition rapidly and 28% of passengers need to use public transportation to avoid this situation.

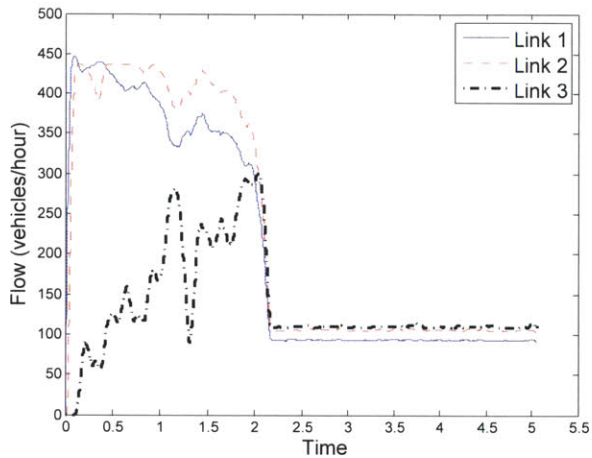
In contrast, at the same level of demand, the UCDP mechanism is able to maintain congestion-free traffic flows on all links. As we can see in Fig. 6-2(e), the densities on three links are always smaller than the corresponding critical densities, which means



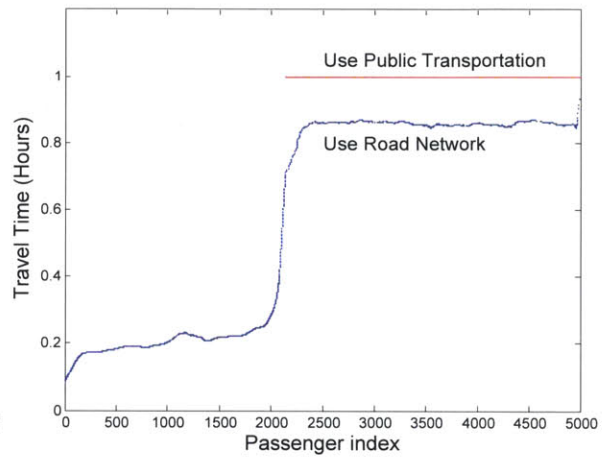
(a) Density (UE).



(b) Link Travel Time (UE).

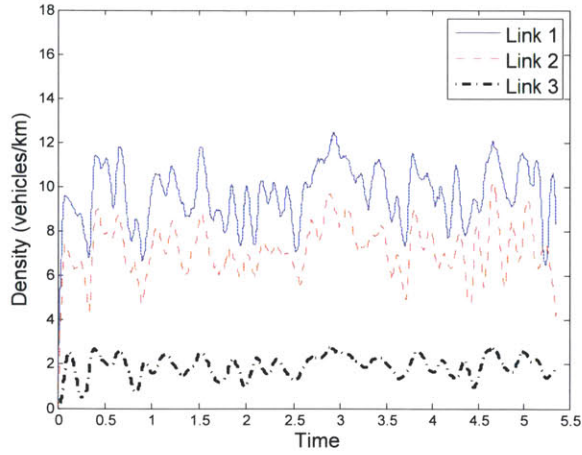


(c) Flow (UE).

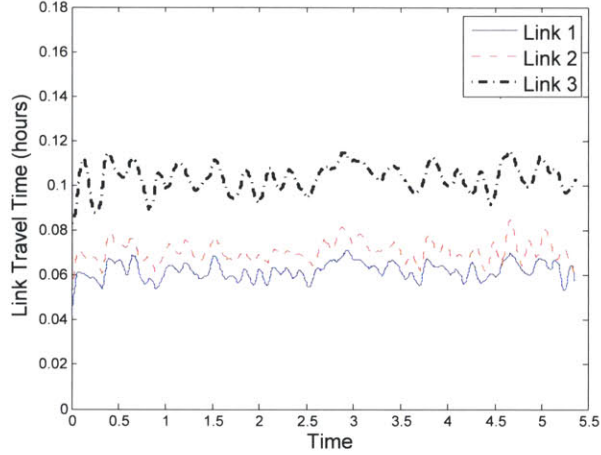


(d) Passengers' Travel Time (UE). Public transport usage: 28%.

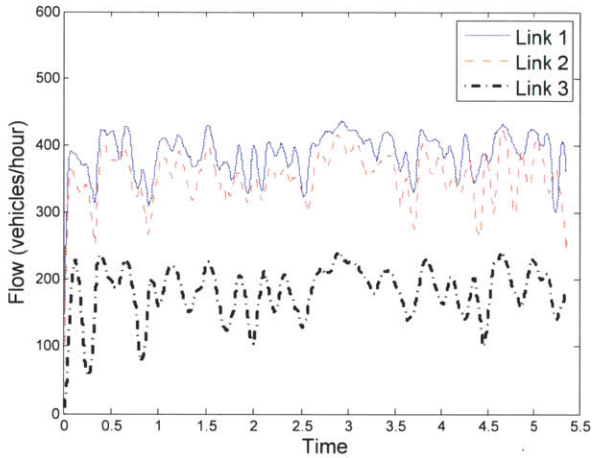
congestion does not happen on these links. This observation is the direct consequence of the constraints considered in the FCSP in Chapter 4. Therefore, the link travel times in Fig. 6-2(f) are stabilized, and the traffic flows in Fig. 6-2(g) are close to the corresponding maximum flows. We report the travel time for each passenger under the UCDP mechanism in Fig. 6-2(h). At equilibrium, we expect that all passengers will report their true maximum tolerated travel time due to the strategy-proof property of the mechanism. The plot indicates that all passengers' preferences are satisfied because their allocated travel times are always less than their maximum tolerated travel times. We also notice that their allocated travel times are generally different



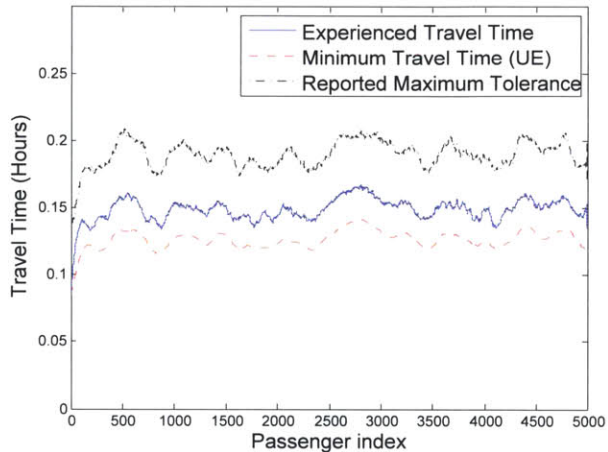
(e) Density (UCDP).



(f) Link Travel Time (UCDP).



(g) Flow (UCDP).



(h) Passengers' Travel Time (UCDP). Public transport usage: 0%

Figure 6-2: Performance of the UCDP mechanism when the demand is 970 v/h.

from the announced minimum times. In addition, the travel time under the UCDP mechanism in Fig. 6-2(h) is about 0.14 hours, which is about 5.7 times faster than the UE performance in Fig. 6-2(d). Overall, the results show that the mechanism is able to maintain congestion-free traffic flows on all links, and thus nobody needs to use public transportation in this case.

When the demand flow rate is increased to 1500 v/h, under UE setting, as shown in Fig. 6-3(a), the network rapidly becomes congested and 77% of passengers need to use public transportation. In contrast, under the UCDP mechanism, Fig. 6-3(b)

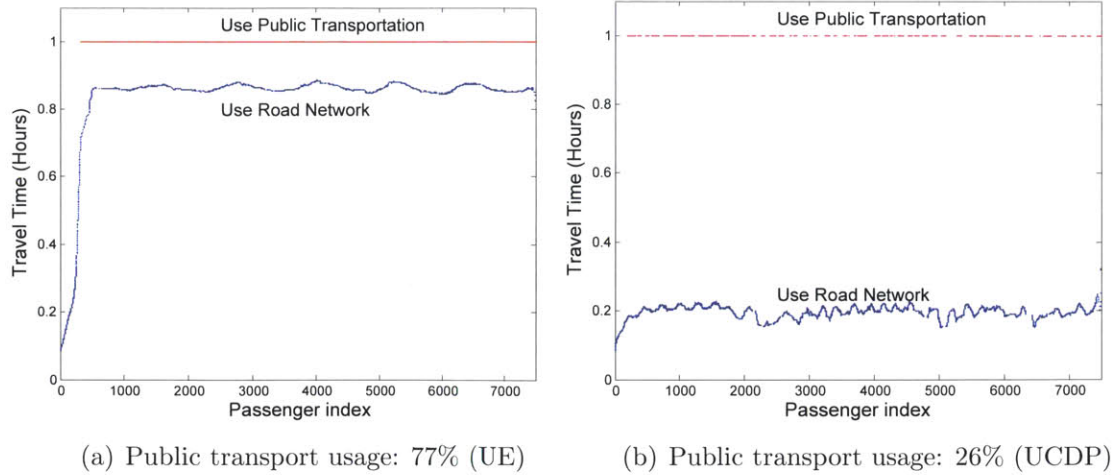


Figure 6-3: Passengers' travel time and public transport usage when the demand is 1500 v/h.

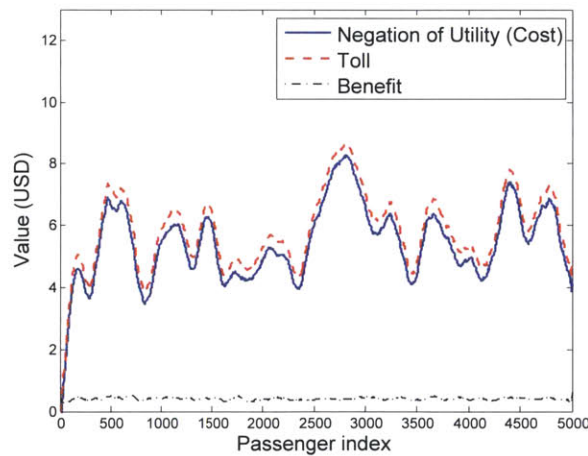
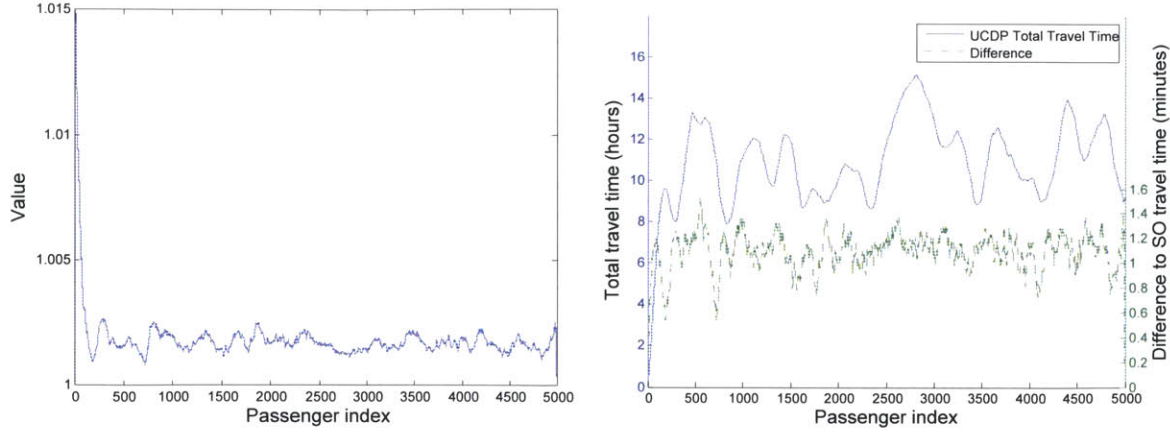


Figure 6-4: Generalized cost and Toll.

shows that passenger travel time is maintained at good level and only 26% of people need to shift to public transportation.

We recall that the mechanism use tolls to incentivize passengers to reveal their true maximum tolerated travel time. The dynamic tolls, generalized costs (negation of utilities), and benefits for passengers in case the demand is 970 v/h are shown in Fig. 6-4. We emphasize that, at equilibrium, all passengers' generalized cost are minimized.

To compare the performance of the UCDP mechanism with SO performance when



(a) UCDP Net Travel Time / SO Travel Time. (b) Numerical values for Fig. 6-5(a).

Figure 6-5: Performance of UCDP mechanism in comparison with the SO performance.

the demand is $970 v/h$, we plot the ratio of the total travel time of all passengers in the network when UCDP is used to the total travel time of all passengers when they are coordinated in a socially optimal way in Fig. 6-5(a). The value of UCDP total travel time and the difference between total travel time in UCDP and SO cases are depicted in Fig. 6-5(b). The two plots indicate that UCDP performance is very close to the SO performance in this experiment. We note that as total demand ($\lambda_1 = 970v/h$) is less than total maximum traffic flows ($1187.5 v/h$) on the three routes, at SO state, there is no congestion. Thus, we emphasize that in UCDP mechanism, we have achieved near SO performance without congestion and at the same time satisfied all passengers' preferences. This observation highlights the game-theoretic aspect of the UCDP mechanism and how the mechanism allocates near SO paths at equilibrium.

6.2 General Network with a Bottleneck

In the second experiment, we demonstrate how the road authority uses the UCDP mechanism to control the traffic flows when road conditions change over time. We consider a general road network, in which link (2,3) is a bottleneck, as shown in Fig. 6-6 with the specifications in Table 6.2. Passengers can go from vertex 0 to 1

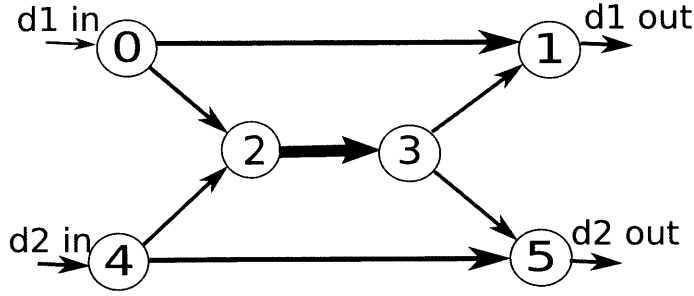


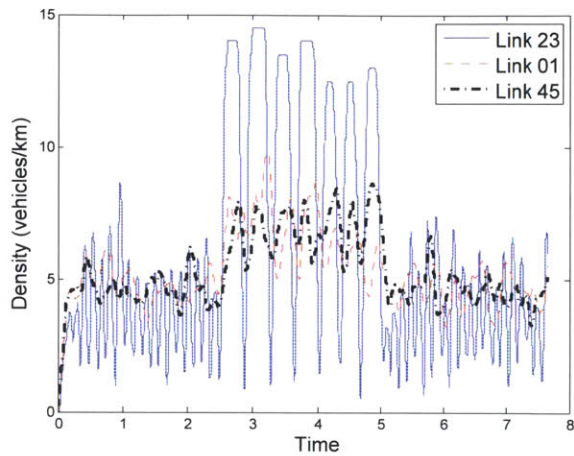
Figure 6-6: General network.

Table 6.2: Network specifications for Fig. 6-6

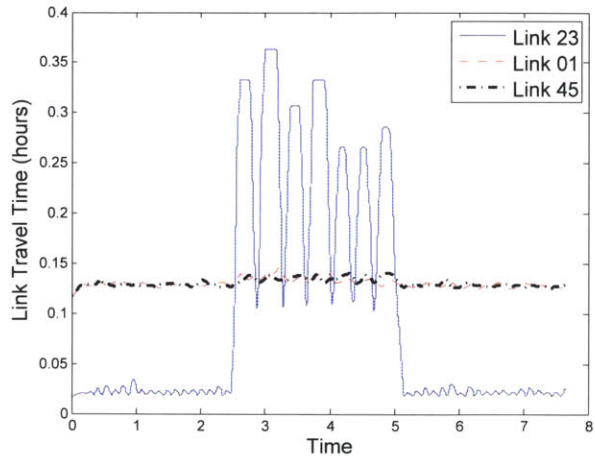
Link	\mathcal{L} (km)	v^{max} (km/h)	K_C (veh./km)	F^{max} (veh./h)
01, 45	7	60.0	25.0	750.0
23	2	120.0 [20.0]	10.0	600.0 [100.0]
others	1.5	30.0	35.0	525.0

or vertex 4 to 5 directly or through the bottleneck, which is shorter. From the table, link (2,3) has the normal maximum speed 120 km/h, but when the link is disrupted, the maximum speed is 20 km/h. There are two demand flows from vertex 0 to vertex 1 and from vertex 4 to vertex 5, each with rate 400 passengers per hour. We assume that value of time for passengers is 10 USD/h. In normal operation, value of time of the road authority is 100 USD/h, but when the link (2,3) is disrupted, the road authority's value of time for link (2,3) is 400 USD/h.

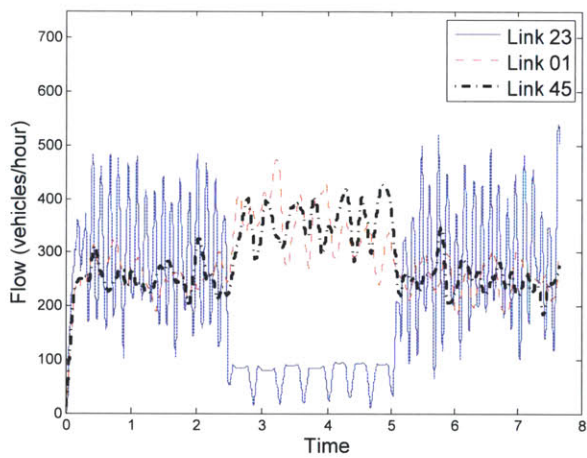
In Fig. 6-7(a)-6-7(c), we show the road network condition over time when the disruption period is from time instant 2.4 to time instant 5.1. As we can see from the density and traffic flow plots, when the disruption occurs, passengers are rerouted to links (0,1) and (4,5). When the link (2,3) is recovered, traffic flows are quickly restored to the normal operation under the UC DP mechanism. Finally, we show the how tolls vary through this disruption in Fig. 6-7(d). As we can see, a majority of passengers is assigned longer paths with lower tolls, and a small fraction of passengers use link (2,3) in their allocated paths with higher tolls. This result reflects the effect of the authority's value of time for link (2,3) under disruption on the distribution of passengers in the road network.



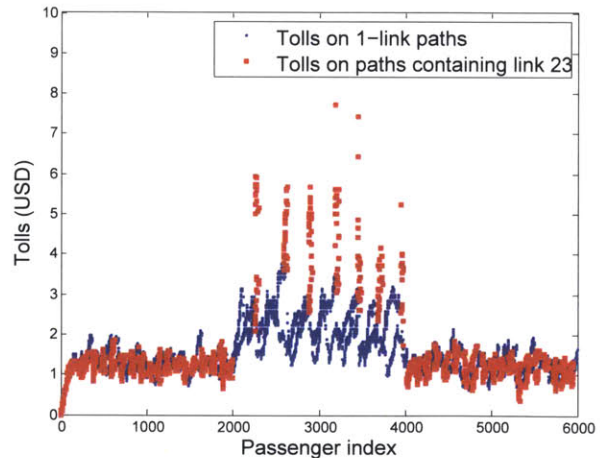
(a) Density.



(b) Link Travel Time.



(c) Flow.



(d) Tolls.

Figure 6-7: Results for the UCDP mechanism on the general network in Fig. 6-6 under traffic disruption. During the disruption period, from time instant 2.4 to time instant 5.1, traffic flows are rerouted to undisrupted paths. The road authority uses his value of time to affect the traffic distribution and tolls.

Chapter 7

Conclusion

In this chapter, we summarize the motivation and objectives of this thesis, the research conducted in this thesis, and the key contributions. We also present several directions for future research.

7.1 Summary

The optimal dynamic traffic routing problem for urban transportation network is considered computationally hard due to its combinatorial nature and its game-theoretic aspects. As transportation networks are complex large-scale systems, designing a real-time traffic routing system that can assign suitable paths to a large number of passengers poses a great challenge to researchers. In addition, selfish and unpredictable travel behavior of drivers make it extremely difficult to enforce them to follow the routing system as well as to foresee potential congested roads. Furthermore, researchers also have concerns about the efficiency of the system as well as the fairness issues arising from heterogeneous classes of passengers. Traditional approaches such as static and dynamic traffic assignment as well as heuristic congestion pricing schemes in real-world deployment do not fully address the above challenges. As a result, we are motivated to explore a novel approach to overcome the computational complexity and satisfy game-theoretic requirements of the problem.

This thesis proposes a novel mechanism called User-Centric Dynamic Pricing (UCDP) based on recent advances in algorithmic mechanism design. The mechanism is applicable, for example, in one-way car rental systems, or mobility-on-demand systems consisting of autonomous cars in future cities. In the mechanism, these services are designed to work with public transportation systems. The mechanism provides a simple protocol for each passenger to interact directly with a road authority and indirectly with other passengers when the passenger wants to join the traffic. The passenger first requests to travel from an origin to a destination. Based on the current network condition, the UCDP mechanism can either suggest him to use public transportation or the road network. In the latter case, the passenger reports his maximum tolerated travel time within the lower bound announced publicly by the road authority. The mechanism then assigns him a path that satisfies his travel time preference while maximizing sequential social surplus and ensuring that congestion does not happen. The mechanism also computes a toll that the passenger should pay to use his assigned path. Unlike other approaches that attempt to compute the path allocation for all drivers at the same time, the UCDP mechanism considers drivers in sequence to reduce the computational complexity of the problem. We have carefully designed algorithms to compute the allocation rule and payment rule so that the mechanism addresses both computational tractability and game-theoretic requirements of the optimal traffic routing problem.

This thesis has the following key contributions. First, the proposed UCDP mechanism is efficient, because it achieves maximum sequential social surplus and prevents congestion. At the same time, the mechanism is user-centric in the sense that it explicitly focuses on each passenger’s travel preference and fairness among individuals. Therefore, the resulting routing system is pleasant and easy to use from passengers’ perspective. More importantly, the mechanism is proved to be incentive compatible, i.e. passengers are always better off to report their true traveling time tolerance. Second, the mechanism can be applied for general networks having multiple OD pairs with dynamic flows of heterogeneous users in a computationally efficient way. The enabling technical idea lies in the new fairness constrained shortest path (FCSP)

problem with a special structure that enables polynomial time computation of path allocation. Third, to the best of our knowledge, this thesis is the first in the literature of transportation that considers general road networks with dynamic flows of heterogeneous users and addresses system efficiency, fairness among passengers as well as computational complexity issues at the same time. These properties of the UCDP mechanism are justified through our analysis and experimental results. Furthermore, viewed from a broader perspective, this novel mechanism equips governments with instruments to achieve sustainable transportation systems by alleviating urban congestion and addressing related social and environmental impacts.

7.2 Future Directions

The extension of this thesis is broad. Several directions for future work are presented here based on our insights gained from this research. More specifically, we would like to investigate the individual rationality aspect of the mechanism, consider multiple travel preferences and design distributed versions of the UCDP mechanism to deal with scalability issues.

First, we want to incorporate individual rationality (IR) in our work so that passengers will naturally join our system. This direction will remove our assumption on IR in this work. Our intended method is based on the following observation. As we have mentioned in the review of VCG mechanism in Section 2.4.3, if the utility of passengers are always non-negative, they do not suffer any loss to join the mechanism. Thus, passengers are willing to join the mechanism. To enable such a scenario, we can offer reward points for passengers who choose to participate in our system. Using results in the theory of repeated games, we aim to show that joining the UCDP mechanism is sustainable with reward points. At the policy making level, passengers' cumulative reward points can be exchanged for tangible benefits such as tax deduction or can be used to establish reputation.

Second, we aim to provide more travel options for passengers to choose from. For examples, besides travel time preference, a passenger can also indicate his max-

imum tolerated toll and specify his route preference. These problems can be formulated as other constrained shortest path problems such as vertex constrained shortest path problems with additional constraints. Although these problems have been well-studied in the literature, designing allocation and payment algorithms for these preferences as well as investigation their complexity is an interesting research direction for future study.

Third, this thesis lays a foundation for future works on designing distributed mechanisms for optimal traffic routing to deal with scalability. In particular, we envision a large-scale network consisting of mobile users who not only share real-time traffic information but also collaborate with each other to achieve social-optimal traffic flows through our future distributed mechanism. Our algorithms to allocate paths and compute tolls in this work is highly suitable for distributed implementation. Furthermore, we observe that most travel requests can be processed by using information in the vicinity of OD pairs. Thus, dividing computation on several computers would significantly reduce computation time for each request. We can further use approximation in these computations to obtain suitable solutions in shorter computing time. Maintaining game-theoretic properties such as incentive compatibility and individual rationality under this approximation is an open problem for future research. Lastly, we would like to extend our mechanism so that it is applicable for road networks governed by not only the public sector but also the private sector. Designing distributed versions of the UC DP mechanism with the participation of the private sector, whose objective is to maximize revenue, poses great challenges for future research.

Bibliography

- [1] U. S. G. A. Office, “Report to the subcommittee on transportation, housing, and urban development and related agencies, committee on appropriations, house of representatives: Traffic congestion, road pricing can help reduce congestion, but equity concerns may grow,” January 2012.
- [2] S. Peeta and A. Ziliaskopoulos, “Foundations of Dynamic Traffic Assignment: The Past, the Present and the Future,” *Networks and Spatial Economics*, vol. 1, no. 3, pp. 233–265, 2001.
- [3] A. S. Schulz and N. E. Stier-Moses, “Efficiency and fairness of system-optimal routing with user constraints,” *Networks*, vol. 48, pp. 223–234, December 2006.
- [4] O. Jahn, R. H. Mhring, A. S. Schulz, and N. E. S. Moses, “System-optimal routing of traffic flows with user constraints in networks with congestion,” *Operations Research*, vol. 53, pp. 4394–2, 2005.
- [5] E. Schmitt and H. Jula, “Vehicle route guidance systems: Classification and comparison,” in *Intelligent Transportation Systems Conference, 2006. ITSC '06. IEEE*, sept. 2006, pp. 242 –247.
- [6] F. A. Chudak, V. D. S. Eleuterio, and Y. Nesterov, “Static Traffic Assignment Problem: A comparison between Beckmann (1956) and Nesterov & de Palma (1998) models,” in *7th Swiss Transport Research Conference*, sept. 2007.
- [7] J. Wardrop, “Some theoretical aspects of road traffic research,” *Proceedings of the Institution of Civil Engineers, Part II*, vol. 1, no. 36, pp. 352–362, 1952.
- [8] O. Drissi-Katouni and A. Hamed-Benchekroun, “A dynamic traffic assignment model and a solution algorithm.” *Transportation Science*, vol. 26, no. 2, pp. 119 – 128, 1992.
- [9] D. K. Merchant and G. L. Nemhauser, “A Model and an Algorithm for the Dynamic Traffic Assignment Problems.” *Transportation Science*, vol. 12, no. 3, p. 183, 1978.
- [10] ———, “Optimality conditions for a dynamic traffic assignment model,” *Transportation Science*, vol. 12, no. 3, pp. 200–207, 1978.

- [11] M. E. Ben-Akiva, “Dynamic network equilibrium research,” *Transportation Research Part A: General*, vol. 19, pp. 429–431, 1985.
- [12] B. N. Janson, “Dynamic traffic assignment for urban road networks,” *Transportation Research Part B: Methodological*, vol. 25, no. 23, pp. 143 – 161, 1991.
- [13] A. Ziliaskopoulos and W. Wardell, “An intermodal optimum path algorithm for multimodal networks with dynamic arc travel times and switching delays,” *European Journal of Operational Research*, vol. 125, no. 3, pp. 486 – 502, 2000.
- [14] M. Goh, “Congestion management and electronic road pricing in Singapore,” *Journal of Transport Geography*, vol. 10, no. 1, pp. 29 – 38, 2002.
- [15] F. T. Seik, “An advanced demand management instrument in urban transport: Electronic road pricing in Singapore,” *Cities*, vol. 17, no. 1, pp. 33 – 45, 2000.
- [16] P. Olszewski and L. Xie, “Modelling the effects of road pricing on traffic in Singapore,” *Transportation Research Part A: Policy and Practice*, vol. 39, no. 79, pp. 755 – 772, 2005.
- [17] W. H. Sandholm, “Evolutionary Implementation and Congestion Pricing,” *Review of Economic Studies*, vol. 69, no. 3, pp. 667–89, July 2002.
- [18] D. Joksimovic, M. C. J. Bliemer, and P. H. L. Bovy, “Optimal Toll Design Problem in Dynamic Traffic Networks with Joint Route and Departure Time Choice,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1923, no. 1, pp. 61–72, 2005.
- [19] O. Chen and M. Ben-Akiva, “Game-Theoretic Formulations of the Interaction between Dynamic Traffic Control and Dynamic Traffic Assignment,” in *Transportation Research Record*, vol. 1617, 1998, pp. 178–188.
- [20] D. Joksimovic, M. C. J. Bliemer, and P. H. L. Bovy, “Road pricing problem including route choice and elastic demand - a game theory approach,” in *TRISTAN V: The Fifth Triennial Symposium on Transportation Analysis, Le Gosier, Guadeloupe*, June 2004.
- [21] N. Nisan and A. Ronen, “Algorithmic Mechanism Design,,” *Games and Economic Behavior*, vol. 35, no. 1-2, pp. 166–196, Apr. 2001.
- [22] R. B. Myerson, “Optimal Auction Design,” *Mathematics of Operations Research*, vol. 6, no. 1, pp. 58–73, 1981.
- [23] R. Lavi, “Computationally efficient approximation mechanisms,” in *Algorithmic Game Theory*. Cambridge Press, 2007, pp. 301–329.
- [24] A. A. Lazar and N. Semret, “The progressive second price auction mechanism for network resource sharing,” in *8th International Symposium on Dynamic Games, Maastricht, The Netherlands*. Citeseer, 1998.

- [25] P. Maillé and B. Tuffin, “The progressive second price mechanism in a stochastic environment,” *Netnomics*, vol. 5, no. 2, pp. 119–147, 2003.
- [26] ———, “Multibid auctions for bandwidth allocation in communication networks,” in *INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 1. IEEE, 2004.
- [27] L. Blumrosen and N. Nisan, “Auctions with severely bounded communication,” in *Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on*. IEEE, 2002, pp. 406–415.
- [28] R. Jain and J. Walrand, “An efficient nash-implementation mechanism for network resource allocation,” *Automatica*, vol. 46, no. 8, pp. 1276–1283, 2010.
- [29] J. Feigenbaum, C. Papadimitriou, R. Sami, and S. Shenker, “A bgp-based mechanism for lowest-cost routing,” *Distributed Computing*, vol. 18, no. 1, pp. 61–72, 2005.
- [30] J. Feigenbaum and S. Shenker, *Distributed algorithmic mechanism design: Recent results and future directions*. September, 2002.
- [31] A. Ozdaglar and R. Srikant, “Incentives and pricing in communication networks,” *Algorithmic Game Theory*, pp. 571–591, 2007.
- [32] F. Kelly, “Charging and Rate Control for Elastic Traffic,” *European Transactions on Telecommunications*, vol. 8, pp. 33–37, 1997.
- [33] F. Kelly, A. Maulloo, and D. Tan, “Rate control in communication networks: shadow prices, proportional fairness and stability,” in *Journal of the Operational Research Society*, vol. 49, 1998.
- [34] H. Yaïche, R. R. Mazumdar, and C. Rosenberg, “A game theoretic framework for bandwidth allocation and pricing in broadband networks,” *IEEE/ACM Trans. Netw.*, vol. 8, pp. 667–678, October 2000.
- [35] J. Feigenbaum, C. Papadimitriou, R. Sami, and S. Shenker, “A BGP-based mechanism for lowest-cost routing,” in *Proceedings of the twenty-first annual symposium on Principles of distributed computing*, ser. PODC ’02. New York, USA: ACM, 2002, pp. 173–182.
- [36] C. Touati, E. Altman, and J. Galtier, “Generalized nash bargaining solution for bandwidth allocation,” *Computer Networks*, vol. 50, no. 17, pp. 3242 – 3263, 2006.
- [37] D. Teodorović, K. Triantis, P. Edara, Y. Zhao, and S. Mladenović, “Auction-based congestion pricing,” *Transportation Planning and Technology*, vol. 31, no. 4, pp. 399–416, 2008.

- [38] K. Wada and T. Akamatsu, “Auction mechanisms for implementing tradable network permit markets,” *Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management)*, vol. 67, no. 3, pp. 376–389, 2011.
- [39] H. Yang and X. Wang, “Managing network mobility with tradable credits,” *Transportation Research Part B: Methodological*, vol. 45, no. 3, pp. 580–594, 2011.
- [40] A. Ozdaglar and R. Srikant, “Incentives and pricing in communications networks,” in *Algorithmic Game Theory*. Cambridge Press, 2007, pp. 571–591.
- [41] M. Ben-Akiva and S. R. Lerman, *Discrete Choice Analysis: Theory and Application to Travel Demand*, 1st ed. The MIT Press, Dec. 1985.
- [42] M. Kilger, “A shadow handler in a video-based real-time traffic monitoring system,” in *Applications of Computer Vision, Proceedings, 1992., IEEE Workshop on*, 1992, pp. 11–18.
- [43] C. Li, K. Ikeuchi, and M. Sakauchi, “Acquisition of traffic information using a video camera with 2d spatio-temporal image transformation technique,” in *Intelligent Transportation Systems, 1999. Proceedings. 1999 IEEE/IEEEJ/JSAI International Conference on*, 1999, pp. 634 –638.
- [44] B. Hoh, T. Iwuchukwu, Q. Jacobson, D. Work, A. M. Bayen, R. Herring, J.-C. Herrera, M. Gruteser, M. Annavaram, and J. Ban, “Enhancing privacy and accuracy in probe vehicle based traffic monitoring via virtual trip lines,” *IEEE Transactions on Mobile Computing*, 2011 2011.
- [45] S. Coleri, S. Y. Cheung, and P. Varaiya, “Sensor networks for monitoring traffic,” in *In 42nd Allerton Conference on Communication, Control and Computing*, 2004.
- [46] M. Fire, D. Kagan, R. Puzis, L. Rokach, and Y. Elovici, “Data mining opportunities in geosocial networks for improving road safety,” in *Electrical & Electronics Engineers in Israel (IEEEI), 2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel*. IEEE, 2012, pp. 1–4.
- [47] W. J. Mitchell, C. E. Borroni-Bird, and L. D. Burns, *Reinventing the Automobile*. MIT Press, 2010.
- [48] Y. Kuwata, J. Teo, G. Fiore, S. Karaman, E. Frazzoli, and J. How, “Real-time motion planning with applications to autonomous urban driving,” *IEEE Trans. on Control Systems Technologies*, vol. 17, no. 5, pp. 1105–1118, 2009.
- [49] K. Treleaven, M. Pavone, and E. Frazzoli, “An asymptotically optimal algorithm for pick-up and delivery problems with application to large-scale transportation systems,” in *IEEE Conf. on Decision and Control*, 2011, submitted.

- [50] L. Immers and S. Logghe, “Traffic flow theory,” *Faculty of Engineering, Department of Civil Engineering, Section Traffic and Infrastructure, Kasteelpark Arenberg*, vol. 40, 2002.
- [51] B. D. Greenshields, J. Bibbins, W. Channing, and H. Miller, “A study of traffic capacity,” in *Highway research board proceedings*, 1935.
- [52] R. D. Kuhne, “Foundations of traffic flow theory i: Greenshields’ legacy—highway traffic,” in *Symposium on the Fundamental Diagram: 75 Years (Greenshields 75 Symposium)*, 2008.
- [53] M. Beckmann, C. B. McGuire, and C. B. Winsten, *Studies in the Economics of Transportation*. Yale University Press, 1956.
- [54] W. W. Mosher, “A capacity-restraint algorithm for assigning flow to a transport network,” *Highway Research Record*, 1963.
- [55] B. V. Martin and M. L. Manheim, “A research program for comparison of traffic assignment techniques,” *Highway Research Record*, no. 88, 1965.
- [56] M. Bruynooghe, A. Gibert, and M. Sakarovitch, “Une methode d’affectation du trafic,” in *Proceedings of the 4th International Symposium on the Theory of Road Traffic*. Bundesminister fur Verkehr, Abt. Strassenbau, Bonn, Germany, 1969, pp. 198–204.
- [57] A. Nagurney, *Network economics: A variational inequality approach*. Springer, 1998, vol. 10.
- [58] D. E. Boyce, H. S. Mahmassani, and A. Nagurney, “A retrospective on beckmann, mcguire and winsten’s studies in the economics of transportation,” *Papers in Regional Science*, vol. 84, no. 1, pp. 85–103, 2005.
- [59] T. Larsson and M. Patriksson, “Side constrained traffic equilibrium modelanalysis, computation and applications,” *Transportation Research Part B: Methodological*, vol. 33, no. 4, pp. 233–264, 1999.
- [60] Y. Sheffi, “Urban transportation networks: equilibrium analysis with mathematical programming methods,” 1985.
- [61] D. Braess, “Ueber ein paradoxen der verkehrsplanung,” *Unternehmensforschung*, vol. 12, pp. 258–268, 1968.
- [62] J. D. Murchland, “Braess’s paradox of traffic flow,” *Transportation Research*, vol. 4, no. 4, pp. 391–394, 1970.
- [63] G. F. Newell, *Traffic flow on transportation networks*, 1980, no. Monograph.

- [64] D. U.S. Department of Transportation, Washington, “Computer programs for urban transportation planning: Planpac/backpac general information manual,” 1977.
- [65] M. L. Manheim and E. R. Ruiter, “Dodotrans i: A decision oriented computer language for analysis of multimodal transportation systems.” *HRB Highway Research Record 314*, pp. 135–163, 1970.
- [66] Y. Sheffi and C. Daganzo, “Hypernetworks and supply-demand equilibrium obtained with disaggregate demand models,” *Transportation Research Record*, no. 673, 1978.
- [67] J. A. Ferland, M. Florian, and C. Achim, “On incremental methods for traffic assignment,” *Transportation Research*, vol. 9, no. 4, pp. 237–239, 1975.
- [68] M. Frank and P. Wolfe, “An algorithm for quadratic programming,” *Naval research logistics quarterly*, vol. 3, no. 1-2, pp. 95–110, 1956.
- [69] J. D. Murchland, “Road traffic distribution in equilibrium,” in *Mathematical Methods in Economic Sciences, Mathematisches Forschungsinstitut, Oberwolfach, West Germany*, 1969.
- [70] L. J. LeBlanc, E. K. Morlok, and W. P. Pierskalla, “An efficient approach to solving the road network equilibrium traffic assignment problem,” *Transportation Research*, vol. 9, no. 5, pp. 309–318, 1975.
- [71] M. Florian and S. Nguyen, “An application and validation of equilibrium trip assignment methods,” *Transportation Science*, vol. 10, no. 4, pp. 374–390, 1976.
- [72] R. Eash, B. N. Janson, and D. Boyce, “Equilibrium trip assignment: advantages and implications for practice,” *Transportation Research Record*, no. 728, 1979.
- [73] Y. N. Fabian A. Chudak, Vania Dos Santos Eleuterio, “Static traffic assignment problem. a comparison between Beckmann (1956) and Nesterov & de Palma (1998) models,” *Swiss Transport Research Conference*, September 2007.
- [74] E. Koutsoupias and C. Papadimitriou, “Worst-case equilibria,” in *STACS 99*. Springer, 1999, pp. 404–413.
- [75] J. K. Ho, “A successive linear optimization approach to the dynamic traffic assignment problem,” *Transportation Science*, vol. 14, no. 4, pp. 295–305, 1980.
- [76] M. Carey, “Optimal time-varying flows on congested networks,” *Operations research*, vol. 35, no. 1, pp. 58–69, 1987.
- [77] J. R. Birge and J. K. Ho, “Optimal flows in stochastic dynamic networks with congestion,” *Operations Research*, vol. 41, no. 1, pp. 203–216, 1993.

- [78] A. Ziliaskopoulos and W. Wardell, “An intermodal optimum path algorithm for multimodal networks with dynamic arc travel times and switching delays,” *European Journal of Operational Research*, vol. 125, no. 3, pp. 486–502, 2000.
- [79] M. Carey and E. Subrahmanian, “An approach to modelling time-varying flows on congested networks,” *Transportation Research Part B: Methodological*, vol. 34, no. 3, pp. 157–183, 2000.
- [80] T. L. Friesz, J. Luque, R. L. Tobin, and B.-W. Wie, “Dynamic network traffic assignment considered as a continuous time optimal control problem,” *Operations Research*, vol. 37, no. 6, pp. 893–901, 1989.
- [81] B. Wie, *Dynamic analysis of user-optimized network flows with elastic travel demand*, 1991, no. 1328.
- [82] B. Ran and T. Shimazaki, “A general model and algorithm for the dynamic traffic assignment problems,” in *Transport Policy, Management and Technology toward 2001: Selected Proceedings of the Fifth World Conference on Transport Research*, vol. 4, 1989.
- [83] ———, “Dynamic user equilibrium traffic assignment for congested transportation networks,” in *Transport Policy, Management and Technology toward 2001: Selected Proceedings of the Fifth World Conference on Transport Research*, vol. 4, 1989.
- [84] B. Ran, D. E. Boyce, and L. J. LeBlanc, “A new class of instantaneous dynamic user-optimal traffic assignment models,” *Operations Research*, vol. 41, no. 1, pp. 192–202, 1993.
- [85] D. E. Boyce, B. Ran, and L. J. Leblanc, “Solving an instantaneous dynamic user-optimal route choice model,” *Transportation Science*, vol. 29, no. 2, pp. 128–142, 1995.
- [86] S. Dafermos, “Traffic equilibrium and variational inequalities,” *Transportation science*, vol. 14, no. 1, pp. 42–54, 1980.
- [87] T. L. Friesz, D. Bernstein, T. E. Smith, R. L. Tobin, and B. Wie, “A variational inequality formulation of the dynamic network user equilibrium problem,” *Operations Research*, vol. 41, no. 1, pp. 179–191, 1993.
- [88] B.-W. Wie, R. L. Tobin, T. L. Friesz, and D. Bernstein, “A discrete time, nested cost operator approach to the dynamic network user equilibrium problem,” *Transportation Science*, vol. 29, no. 1, pp. 79–92, 1995.
- [89] B. Ran and D. E. Boyce, “A link-based variational inequality formulation of ideal dynamic user-optimal route choice problem,” *Transportation Research Part C: Emerging Technologies*, vol. 4, no. 1, pp. 1–12, 1996.

- [90] B. Ran, R. W. Hall, and D. E. Boyce, “A link-based variational inequality model for dynamic departure time/route choice,” *Transportation Research Part B: Methodological*, vol. 30, no. 1, pp. 31–46, 1996.
- [91] H.-K. Chen and C.-F. Hsueh, “A model and an algorithm for the dynamic user-optimal route choice problem,” *Transportation Research Part B: Methodological*, vol. 32, no. 3, pp. 219–234, 1998.
- [92] M. Van Aerde and S. Yagar, “Dynamic integrated freeway/traffic signal networks: a routing-based modelling approach,” *Transportation Research Part A: General*, vol. 22, no. 6, pp. 445–453, 1988.
- [93] ———, “Dynamic integrated freeway/traffic signal networks: Problems and proposed solutions,” *Transportation Research Part A: General*, vol. 22, no. 6, pp. 435–443, 1988.
- [94] J. Barceló, J. Ferrer, and R. Grau, “Aimsun2 and the getram simulation environment,” in *Proc. 13th EURO Conference*, 1994.
- [95] R. Jayakrishnan, H. S. Mahmassani, and T.-Y. Hu, “An evaluation tool for advanced traffic information and management systems in urban networks,” *Transportation Research Part C: Emerging Technologies*, vol. 2, no. 3, pp. 129–147, 1994.
- [96] H. Mahmassani, T. Hu, and R. Jayakrishnan, “Dynamic traffic assignment and simulation for advanced network informatics (dynamart),” *Urban traffic networks: Dynamic flow modeling and control*. Springer, Berlin/New York, 1995.
- [97] M. Ben-Akiva, M. Bierlaire, H. Koutsopoulos, and R. Mishalani, “Dynamit: a simulation-based system for traffic prediction,” in *DACCORS Short Term Forecasting Workshop, The Netherlands*. Citeseer, 1998.
- [98] A. Pigou, “The economics of welfare,” *McMillan&Co., London*, 1920.
- [99] R. Linsey and E. Verhoef, “Traffic congestion and congestion pricing,” *Handbook of Transport Systems and Traffic Control (eds., Button K.J. and Hensher D.A.)*, pp. 77–105, 2001.
- [100] A. A. Walters, “The theory and measurement of private and social cost of highway congestion,” *Econometrica: Journal of the Econometric Society*, pp. 676–699, 1961.
- [101] W. S. Vickrey, “Congestion theory and transport investment,” *The American Economic Review*, vol. 59, no. 2, pp. 251–260, 1969.
- [102] R. Arnott and K. Small, “The economics of traffic congestion,” *American Scientist*, vol. 82, no. 5, pp. 446–455, 1994.

- [103] T. Tsekeris and S. Voß, “Design and evaluation of road pricing: state-of-the-art and methodological advances,” *Netnomics*, vol. 10, no. 1, pp. 5–52, 2009.
- [104] J. R. Meyer, J. A. Gómez-Ibáñez, W. W. B. Tye, and C. Winston, *Essays in transportation economics and policy: a handbook in honor of John R. Meyer*. Brookings Institution Press, 1999.
- [105] G. Santos, W. W. Li, and W. T. Koh, “9. transport policies in singapore,” *Research in Transportation Economics*, vol. 9, pp. 209–235, 2004.
- [106] A. de Palma and R. Lindsey, “Traffic congestion pricing methodologies and technologies,” *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1377–1399, 2011.
- [107] (2013, July) Toll schedules - 91 express lanes. [Online]. Available: www.91expresslanes.com/schedules.asp
- [108] A. de Palma, “A game-theoretic approach to the analysis of simple congested networks,” *The American Economic Review*, vol. 82, no. 2, pp. 494–500, 1992.
- [109] O. J. Chen and M. E. Ben-Akiva, “Game-theoretic formulations of interaction between dynamic traffic control and dynamic traffic assignment,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1617, no. 1, pp. 179–188, 1998.
- [110] D. Joksimovic, M. C. Bliemer, and P. H. Bovy, “Different policy objectives of the road-pricing problem: a game-theoretic approach,” *Pricing in road transport: a multi-disciplinary perspective*, pp. 151–169, 2008.
- [111] D. Fudenberg and J. Tirole, *Game theory*. 1991. MIT Press, 1991.
- [112] C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou, “The complexity of computing a nash equilibrium,” *SIAM Journal on Computing*, vol. 39, no. 1, pp. 195–259, 2009.
- [113] N. Nisan, “Introduction to Mechanism Design (for Computer Scientists),” in *Algorithmic Game Theory*. Cambridge Press, 2007, pp. 209–241.
- [114] R. B. Myerson and M. A. Satterthwaite, “Efficient mechanisms for bilateral trading,” *Journal of economic theory*, vol. 29, no. 2, pp. 265–281, 1983.
- [115] T. Akamatsu, “Tradable network permits: A new scheme for the most efficient use of network capacity,” *submitted to Transportation Science*, 2007.
- [116] U. Springer, “The market for tradable ghg permits under the kyoto protocol: a survey of model studies,” *Energy Economics*, vol. 25, no. 5, pp. 527–551, 2003.
- [117] V. William, “Counterspeculation, Auctions, and Competitive Sealed Tenders,” *The Journal of Finance*, vol. 16, no. 1, pp. 8–37, 1961.

- [118] E. H. Clarke, “Multipart pricing of public goods,” *Public choice*, vol. 11, no. 1, pp. 17–33, 1971.
- [119] T. Groves, “Incentives in Teams,” *Econometrica*, vol. 41, no. 4, pp. 617–631, 1973.
- [120] N. Nisan and A. Ronen, “Computationally feasible vcg mechanisms,” in *ACM Conference on Electronic Commerce*, 2000, pp. 242–252.
- [121] A. Archer and É. Tardos, “Truthful mechanisms for one-parameter agents,” in *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on.* IEEE, 2001, pp. 482–491.
- [122] M. H. Rothkopf, A. Pekec, and R. M. Harstad, “Computationally manageable combinatorial auctions,” Tech. Rep., 1995.
- [123] G. Gallo and S. Pallottino, “Shortest path algorithms,” *Annals of Operations Research*, vol. 13, no. 1, pp. 1–79, 1988.
- [124] E. W. Dijkstra, “A note on two problems in connexion with graphs,” *Numerische mathematik*, vol. 1, no. 1, pp. 269–271, 1959.
- [125] R. Bellman, “On a Routing Problem,” *Quarterly of Applied Mathematics*, vol. 16, pp. 87–90, 1958.
- [126] L. R. Ford, “Network flow theory.” 1956.
- [127] D. Ford and D. R. Fulkerson, *Flows in networks*. Princeton university press, 1962.
- [128] U. Pape, “Implementation and efficiency of moore-algorithms for the shortest route problem,” *Mathematical Programming*, vol. 7, no. 1, pp. 212–222, 1974.
- [129] D. P. Bertsekas, “A simple and fast label correcting algorithm for shortest paths,” *Networks*, vol. 23, no. 8, pp. 703–709, 1993.
- [130] J. Hao and G. Kocur, “An implementation of a shortest augmenting path algorithm for the assignment problem,” *Network flows and matching: first DIMACS implementation challenge, Series in DIMACS*, vol. 12, pp. 453–468, 1993.
- [131] J. Beasley and N. Christofides, “An algorithm for the resource constrained shortest path problem,” *Networks*, vol. 19, no. 4, pp. 379–394, 1989.
- [132] M. Desrochers and F. Soumis, “A generalized permanent labelling algorithm for the shortest path problem with time windows,” *INFOR*, vol. 26, no. 3, pp. 191–212, 1988.
- [133] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness (Series of Books in the Mathematical Sciences)*, 1st ed. W. H. Freeman & Co Ltd, Jan. 1979.

- [134] R. C. Larson and A. R. Odoni, *Urban operations research*, 1981, no. Monograph.
- [135] K. Holmberg and D. Yuan, “A multicommodity network-flow problem with side constraints on paths solved by column generation,” *INFORMS Journal on Computing*, vol. 15, no. 1, pp. 42–57, Winter 2003.