SELFISH ROUTING WITH USER PREFERENCES

GAME THEORETICAL STUDY

ON

SELFISH ROUTING WITH USER PREFERENCES

By

HAO XIA, B.E.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Science

McMaster University

MASTER OF SCIENCE (2008) McMaster University
COMPUTING AND SOFTWARE Hamilton, Ontario

TITLE: Game Theoretical Study on Selfish Routing with User Preferences

AUTHOR: Hao Xia, B.E. (Tsinghua University)

SUPERVISOR: Dr. George Karakostas

NUMBER OF PAGES: x, 63

Abstract

We study selfish routing with user-specific preferences. The selfish routing model captures the selfish behaviour of users in a transportation system. Each user tries to minimize her own travel latency by choosing the shortest route, i.e. the route with the smallest latency, without taking other users' welfare into consideration. In this model, users are assumed infinitesimal, in the sense that the impact of the behaviour of any single user to the network is negligible. Under certain constraints, a steady state is known to exist, where no user has the incentive to deviate from her current route. This state is referred to as a traffic equilibrium.

We extend the traditional selfish routing model by incorporating user preferences. In traditional selfish routing, one assumes that users make their routing decisions merely based on path latencies, and furthermore all users perceive the same latency on any single edge. We observe that in reality users may have their personal preferences on the routes to travel on, e.g., some may be accustomed to certain routes and feel unwilling to try out new ones; some may enjoy the wonderful views available on some routes more than their care for the slightly longer travel time; some may be limited to certain routes because they prefer public transportations, and so on. This observation motivates our work. We introduce to the model a set of edgewise user-specific preferences which come as inherent properties of each user. The disutility of using each edge, which is the basis of a user's routing decision, is now a function of latency and user preference.

In this work several equilibria related aspects of the extended model are studied. Since we are working with infinitesimal users, the distinct combinations of edge

preferences among all users might be either finite or infinite. There is qualitative difference between these two cases in terms of analysis, and both cases are discussed. We show existence and uniqueness of equilibria, give an upper bound on the price of anarchy, and study how taxation can help in this setting.

Acknowledgements

First and foremost, I would like to thank my supervisor George Karakostas, without whom this thesis will not be possible. He inspired my research interest in theoretical computer science, and has been a great tutor. Our frequent meetings were my main source of new ideas, and he has always been helpful in pointing out where I should look at next.

I would also like to give my thanks to Michael Soltys, who taught an excellent course on complexity theory during my first term at McMaster University. I never knew that complexity theory could be so interesting.

I also want to thank my friends for their friendship and companionship. Life will be a misery without them.

Last but not the least, I am indebted to my parents, who have always been a strong support to me. Their love and encouragement helped me out of those hard times.

Hamilton, Ontario, Canada Hao Xia

April, 2008

Contents

List of Figures

Chapter 1

Introduction

1.1 Selfish Routing

Almost everybody travels on a day-to-day basis. People start off from different places, and bear different destinations in mind. Normally there will be more than one route that can lead a traveller to her destination, and it is up to her decision which route to travel on. Transportation resources, most notably roads, are shared among all travellers. Same as in other resource sharing systems, there is the phenomenon called *congestion*, that is when a lot of people are travelling on the same road, it usually takes a long time to get through — almost everyone has experienced being caught in a traffic jam. From a bird's-eye view, the freedom of choice on the route to travel on for everybody induces a variety of travel patterns. However, there is often an intention behind each traveller making her decision, which most probably is to get to her destination as fast as possible. When it comes to this point, people are very likely to be *selfish*, in the sense that they do not care about how their choices will influence other fellow travellers.

This commonplace experience is where the model of *selfish routing* originates from. The system is formulated as a *directed network* with a given set of node pairs called the *origin-destination* (O-D) pairs. With each O-D pair, a fixed *demand* is associated, representing the amount of users to be routed correspondingly per unit of time. An O-D pair together with its associated demand is usually referred to as a *commodity*. Each path in this network is prescribed with a *latency function* that quantifies the common congestion to be experienced by all users using the path according to the current routing pattern. Each user behaves selfishly, with the objective of minimizing travel latency for herself without taking other users' welfare into consideration. In this model, users are assumed *infinitesimal*, in the sense that the impact of the behaviour of any single user to the network is negligible. Under certain assumptions, a steady state exists, in which no user has the incentive to deviate from her current path. This steady state is then referred to as an *equilibrium*. Although selfish routing originated in transportation systems, it is also possible to adapt it to routing on data networks, e.g. the Internet.

In this work, since we will be extending the selfish routing model, we shall sometimes refer to it as *traditional* selfish routing in order to differentiate it from our extended model.

1.2 User Preferences

In the traditional selfish routing model described above, an implicit assumption is that users make their routing decisions merely based on path latencies, and furthermore all users perceive the same latency on any single path. This means that every traveller judges the routes using exactly the same standard. We observe that this is too strong an assumption in many cases, if not unrealistic, as it rules out the *heterogeneity* among different users. Users are likely to have their personal preferences on the routes to travel on. For example, some users may be accustomed to certain routes and feel unwilling to try out new ones; some may enjoy the view available on some routes more than their care for the slightly longer travel time; some may be limited to certain routes because they prefer public transportations, and so on. This inherent heterogeneity among individual users, on an abstract level, can be summarized by *nonuniform perceptions* of path latencies in the network. And this motivated our work.

We extend traditional selfish routing with user-specific *preferences* which come as inherent properties of each user and are assumed constant. In our model, users are as selfish are they always are, but their decisions will involve both path latencies and personal preferences. It is interesting to point out that, during the writing of this thesis, we noticed the recent work of [MMMT07] in which the authors studied congestion games with player-specific constants, the idea of which is very similar to our user preferences. We see that their network congestion games are equivalent to selfish routings with *atomic* players, as opposed to non-atomic players or infinitesimal users, which is our case. In spite of the difference in types of games, their work shares much resemblance in motivation with ours, which on the other hand further supports the practical importance of our extension.

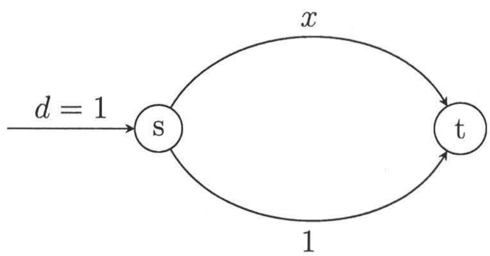

Figure 1.1: Pigou's example

1.3 Previous Work

It has long been known that, in general, such selfish behaviour may not lead to optimal social welfare which is measured by the total latency experienced by all users on the network. The best example to illustrate this fact, as well as to obtain a first feeling of selfish routing, is Pigou's example, which essentially originated from [Pig20] in 1920. In Pigou's example, shown in Figure 1.1, the network consists of only two nodes and two parallel links. One of the links is prescribed with latency function x, i.e. linear to the amount of users using it; the other link bears constant latency 1 no matter how much traffic is put upon it. There is a total demand 1 of infinitesimal users to be routed from s to t. Then intuitively one should be able to predict how selfish users will travel through this network. All of them will choose the linear link which looks like a "safe shot", as the worst latency they can ever have is no worse than going through the constant link. Then if we look at the total latency experienced by all users, pure selfishness gives us 1, which is by multiplying the per-person latency with the total amount of users. Now imagine there is some central coordinator who regulates traffic in a half-half fashion, i.e., both links get half of the users travelling through them. As a result, the link latencies will be 1/2 and 1 respectively, and half of the users will be happier than before, while the other half will just remain unchanged. The total latency in the coordinated case is 3/4, which is in fact the best result one can get in this example. So 1 versus 3/4, this is the loss of social welfare caused by selfishness (anarchy).

It is natural to measure this inefficiency of selfishness by the ratio of the worst-case selfish social cost to the optimal social cost. This ratio is usually referred to as the *price of anarchy* [Pap01]. For instance, it is 4/3 in Pigou's example. The price of anarchy in networks with general topologies was first theoretically studied by Roughgarden and Tardos [RT02], where the authors produced a tight upper bound on it for networks with linear latency functions. Interestingly, the upper bound is 4/3, implying that Pigou's example is the worst case among all networks with linear latency functions. Perakis [Per04] generalized it to networks where latency functions are non-separable and asymmetric, and provided an alternative proof of the bound in [RT02]. Correa, Schulz, and Stier-Moses [CSSM04] then proposed the notion of β-function, as an extension to the anarchy value α in [Rou03], to categorize latency functions, thus generalizing the result to general latency functions. They also gave an even simpler proof of the 4/3 bound for linear latencies.

With this inefficiency of selfishness, it arises as a natural quest for system designers to think of means to cope with it. While it is possible to tackle this problem with modifications to network topologies or latency functions, most of the time it is more realistic to deal with selfishness by means of *taxation*. Faced with extra disutility taxes on each choice of route, users will have to trade off between latencies and taxes when making their routing decisions. In case users are homogeneous in the latency-tax tradeoff, the classical *marginal cost taxation* due to [Pig20], i.e. the tax of using a route is equal to the additional congestion induced to all other users, is known to be capable of inducing equilibria that minimize total latency [BMW56]. The taxation for heterogeneous users, i.e. users with heterogeneous trade-offs between tax and latency, appeared much later in literature. Yang and Huang [YH04] studied this case and showed the existence of optimal taxes that induce equilibria that optimize the system performance. Similar results appeared also in [CDR03, KK04a].

Another direction of research on selfish routing is to extend or modify the model to accommodate more general or more specific problems. Aashtiani and Magnanti [AM81] proved existence and uniqueness of equilibria on a model with non-separable latency functions and elastic demands, that is the demand of each commodity is no longer a constant but depends on the cost of routing this commodity. [CSSM04] considered the case where edges of the network have capacities, and bounded the ratio of the *best* equilibrium social cost to the optimal social cost, as the *worst* equilibrium in the capacitated model can be arbitrarily bad. In [KKVX07] the authors extended selfish routing by allowing a fraction of oblivious users, who make routing decisions based only on static characteristics of the network and are oblivious to congestion. Daganzo and Sheffi [DS77] introduced errors to user perceptions of latencies, and characterized equilibria in a stochastic sense. In the model of [MNS04], user strategies are edge based rather than path based, in which case by fixing a strategy, a user is not bound to travel on a certain route. [FOV07] assigns priorities to each user where users with higher priorities experience less delay when traversing the same edge as compared to users with lower priorities.

1.4 Our Contribution

In this work, we formally introduce edgewise user preferences as a given input to the traditional selfish routing setting. We model them using a *penalty function* t that maps each user to a nonnegative edgewise penalty value. The smaller the penalty value, the more preferable the particular edge to the user. Edgewise penalties contribute to edge disutilities as an additive term, that is, the disutility of using edge e for user r is the summation of latency on e due to congestion, and the user-specific penalty value $t_e(r)$ of e. Since we are working with infinitesimal users, the range of t may be a *finite* set or an *infinite* set. In other words, there may be finitely or infinitely many different penalties among all users. There is a qualitative difference between these two cases in terms of analysis, and both cases are discussed in this thesis. Our contributions include the following:

- We extend the traditional selfish routing by incorporating user-specific preferences. We prove the existence of equilibria in the extended model as a direct application of [Sch73, Theorem 2]. We prove the uniqueness of equilibria, that is, for a given problem setting, all equilibria induce identical edge latencies. Notably, our uniqueness result implies the validity of the conjecture made in [CDR03] after Proposition 2.5.

- We show that the price of anarchy for traditional selfish routing given in [CSSM04] remains valid in our extended model. This shows that introducing nonnegative user-specific penalties does not degrade the efficiency of equilibria.

- We study the existence of taxes that drive users into desirable flow patterns. We consider two separate objectives: to minimize the *social cost*, i.e. cost including both latencies and penalties, and to minimize the *total latency.* When users have homogeneous tax sensitivities, we show that marginal cost taxation suffices to minimize the social cost. We also show that no optimal social cost tax exists when users have heterogeneous tax sensitivities in general by a counter example. For optimal total latency, we observe that the proving techniques in [KK04a] as well as their results apply in our model.

1.5 Organization

In Chapter 2 we give some basics on selfish routing and present the formal formulation of our extended model. Chapter 3 discusses properties of equilibria in our model, including existence, uniqueness, and the price of anarchy. In Chapter 4 we study the existence of two kinds of optimal taxes: one that minimizes social cost, and one that minimizes total latency. Some problems and subtleties unexplored in the main body of the thesis will be discussed with more details in Chapter 5.

Chapter 2

Preliminaries

In this chapter, preliminary knowledge required in the rest of this thesis is reviewed. In Section 2.1 we go through a few basic concepts in game theory and give details about the model of traditional selfish routing. We then formally introduce user preferences by means of penalty function in Section 2.2. In Section 2.3 our extended model is formally introduced.

2.1 Selfish Routing Basics

Before we actually move onto selfish routing, a few words about some fundamental concepts in game theory might be helpful in grasping the basic ideas of selfish routing. However we do not aim at precise definitions or rigorous proofs, for which one should refer to textbooks on game theory, e.g. [FT91].

A *game* in *normal form* consists of a set of *players*, a set of available *strategies* for each player, and a *payoff function* that maps strategy profiles to payoffs for each player. A *strategy profile* is a set of strategies with one strategy from each player's available strategy set. In other words, a strategy profile denotes the chosen strategy of each player. We will consider *non-cooperative* games, in which players are self centric and play to maximize their own payoffs while paying no attention to other players' payoffs. A (pure) equilibrium, or Nash equilibrium, is a strategy profile such that no player can achieve a larger payoff by unilaterally changing her strategy. Thus at equilibrium, assuming all players are rational, no player would have the incentive to change her chosen strategy, hence resulting in a steady state. A priori, no pure equilibrium may exist in a game.

Selfish routing is an example of a non-cooperative game. The setting is a network, formulated as a finite directed graph. The set of users[1] is an infinite set but with a finite measure representing the total amount of users using the network. Probably the best way to comprehend this is by thinking of points on a closed real interval. Users are grouped by their origins and destinations, i.e. where they come from and want to travel to, on the network. A strategy is a simple path on the network, and the set of available strategies for a user is the set of simple paths connecting her origin and destination, which we always assume is nonempty. A strategy profile will induce a path[2] flow on the network, characterized by the amount of users using each path. The disutility, i.e. the negative payoff, of a user is the latency on her chosen path, given the path flow induced by current strategy profile. If path latency functions are continuous, then one can observe Wardrop's principle [War52]:

> At equilibrium, for each origin-destination pair, the latencies of all used paths are equal, and less than or equal to those of unused paths.

[1]In selfish routing, people often refer to players as users.
[2]We shall use *path* to refer to *simple path* if not stated otherwise.

In traditional selfish routing, we do not distinguish between individual users with the same origin and destination, and we are only concerned with macroscopic variables such as flows. Thus there is the concept of *equilibrium flow*, which is the induced path flow of a Nash equilibrium of the game.

In the aforementioned setup, we stated only that path latencies are functions of path flows. This is quite flexible in general which often implies complexity in the analysis. There is a more restrictive version, called the *additive model*, which is widely adopted in modeling real life problems. In the additive model, the latency of an edge is a function of path flows, and the latency of a path is given by the summation of the latency of its edges. If we use $l_p(f)$ and $l_e(f)$ to represent the latency on path p and on edge e respectively given a flow f, then the additive model can be expressed as $l_p(f) = \sum_{e \in p} l_e(f)$. If we further require that the latency of an edge depends only on the flow through that edge, then we have the case which is called *separable* latency functions. Usually when we refer to selfish routing, we mean selfish routing with separable latency functions.

The proof for the existence of equilibria in selfish routing is not trivial though. If the latency functions are continuous, nondecreasing, and separable, then the existence of equilibria is achieved by formulating equilibrium flows as solutions to a convex program [BMW56]. With non-separable latency functions, if they are continuous and nonnegative, then formulating equilibrium flows as solutions to a nonlinear complementarity problem and applying Brouwer's fixed-point theorem can show the existence of equilibria [AM81].

2.2 Infinitesimal Users & Preferences

Here we formalize the concept of *infinitesimal users.* The set of infinitesimal users is modeled using a closed real interval, and without loss of generality we use $[0, 1]$, endowed with Lebesgue measure λ. Then whenever we want to quantify the *amount* of a subset of users, λ will be used as the measure. For readers not familiar with the Lebesgue measure, it can be intuitively thought of as a way of measuring the volume or length of a set. One should note that this representation naturally implies that users are unweighted.

To model user preferences we use a penalty function t on domain $[0, 1]$. The codomain of t is the set of edgewise penalty assignments which we will introduce formally in Section 2.3. As a preview, one important thing about the penalty function is that it may take on either finitely or infinitely many distinct values, and in the latter case it means users can not be grouped into a finite number of groups according to their preferences. This is the key point that distinguishes finite and infinite instances in the rest of this work.

2.3 Our Model

General Model. Let $G = (V, E)$ be a finite digraph (possibly with parallel edges but no self loops) with P denoting the set of all simple paths in G. We shall use P_w for any node pair $w \in V \times V$ to denote the set of all simple paths connecting w in G.

Let the real interval $R = [0, 1]$ endowed with Lebesgue measure λ denote the set of all *users* traveling through the network. Each user $r \in R$ is associated with a pair of nodes, representing his origin and destination. This association is given by the *O-D function* $d : R \to V \times V$. Let D be the range of d, i.e., D is the set of all O-D pairs; then we require for each O-D pair $w \in D$, that P_w is nonempty. We also denote that $P_r = P_{d(r)}$, which is the set of paths available for user r, for ease of exposition.

A *flow* in our general model is a measurable function $f : R \to P$, i.e. an assignment of users to paths. Notice that the definition of flow in the general model is different than the traditional sense. A flow f is said to be *feasible* if it satisfies the demands, that is, if a.e.[3] $f(r) \in P_r$. Denote the set of all *feasible flows* by F. For each path $p \in P$, the *path flow* (with respect to flow f) is defined as $f_p = \lambda(\{r \in R : f(r) = p\})$, representing the amount of users using path p in flow f. For each edge $e \in E$, denote the *edge flow* by $f_e = \sum_{p \ni e} f_p$.

To model congestion effects, each edge e is associated with a real-valued *latency function* l_e, i.e., $l_e(f_e)$ is the common latency to be experienced by all users using edge e when a.e. user travels according to f. Most of the time we will be dealing with l_es that are *continuous* and *nondecreasing* in this thesis. We also assume the additive model where *path latencies* are given by $l_p(f) = \sum_{e \in p} l_e(f_e)$ for each path p.

Each user can have his personal preferences on edges. This is captured by the edge *penalty(dislike) function* $t : R \to \mathbb{R}_{\geq 0}^{|E|}$ where $t_e(r)$ measures the penalty of using edge e for user r. Conceivably the most preferred edge for a user will be assigned the smallest penalty value. Also we let $t_p(r) = \sum_{e \in p} t_e(r)$ denote the penalty of using path p for user r. We will assume that t is measurable by default.

[3]The acronym *a.e.* stands for *almost every* or *almost everywhere*. This is a terminology used with Lebesgue measures to denote that the subset of objects being quantified that do not satisfy the subsequent statement is a *null set*, i.e. a set with zero measure.

We shall call the quadruple (G, d, l, t) a *(general) instance* of our problem. The rest of this section will be based on some instance (G, d, l, t) if we do not make explicit declarations.

The *disutility* a user sees comprises two parts: the *latency* imposed on edges and incurred by congestion, and the *penalty* as a result of her personal preference. Formally, the disutility of using edge e for user r when a.e. user travels according to flow f is defined as $u_e(r, f) = l_e(f_e) + t_e(r)$. The path disutility is given by $u_p(r, f) = \sum_{e \in p} u_e(r, f) = l_p(f) + t_p(r)$ for each path p. Also let $u(r, f) = u_{f(r)}(r, f)$ for simplicity in symbolism, where $f(r)$ is the path chosen by user r in flow f. Then $u(r, f)$ is the traveling disutility of user r in flow f.

The *social cost* c induced by flow f is the total disutility of all users when almost everyone travels according to f. Formally the definition of the social cost is $c(f) = \int_R u(r, f) d\lambda$. Furthermore let us define an *auxiliary cost function* $c_{f^1}(f^2) = \int_R u_{f^2(r)}(r, f^1) d\lambda$ for any flows f^1 and f^2. $c_{f^1}(f^2)$ can be interpreted as the total disutility induced by f^2 on a constant latency network where for each edge e, the edge latency is given by $l_e(f_e^1)$. Note that we have $c(f) = c_f(f)$ by definition.

We say that a flow f^* is at *equilibrium* iff it is feasible and, for a.e. user r,

$$u(r, f^*) \leq u_p(r, f^*), \quad \forall p \in P_r. \tag{EC}$$

Notice that this *equilibrium condition* is equivalent to saying

$$c(f^*) \leq c_{f^*}(f), \quad \forall f \in F. \tag{EC'}$$

Going from (EC) to (EC') is a simple result of the monotonicity of Lebesgue integration and the fact that null sets can be ignored when doing Lebesgue integration. The reverse direction can be seen by constructing a new flow that only varies with f^* in those users that are not "satisfied", thus arriving at a contradiction to (EC') inequality. Due to the equivalence between the two formulations, we will simply use (EC) to refer to either formulation in the rest of this thesis.

Discrete Model. Consider some instance (G, d, l, t) where the range of t is a finite set. In other words, there are only finitely many different penalty assignments among all users. We shall say that such t is finite and (G, d, l, t) is a *finite instance.* We will construct a so called *discrete instance* with respect to this finite instance, which will be a traditional selfish routing game with *nonseparable* latency functions.

Let $\mathcal{T} = \{t(r) : r \in R\}$ be the set of all *penalty assignments.* Note again that $\mathcal{T}$ is a finite set. For each $\tau \in \mathcal{T}$ we let $G^\tau = (V^\tau, E^\tau)$ be a copy of G labeled with τ. Denote the set of all labeled nodes as $\mathcal{V} = \bigcup_{\tau\in\mathcal{T}} V^\tau$ and the set of all labeled edges as $\mathcal{E} = \bigcup_{\tau\in\mathcal{T}} E^\tau$. Hence we have a huge yet finite digraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ which is the result of aggregating several Gs with different labels together. Notice that each G^τ is an *isolated component* of $\mathcal{G}$ since we do not add extra edges to connect them. Thus the set of all simple paths in $\mathcal{G}$ is simply given by $\mathcal{P} = \bigcup_{\tau\in\mathcal{T}} P^\tau$. As usual, we use $\mathcal{P}_\omega$ to denote the set of simple paths connecting ω in $\mathcal{G}$ for each node pair $\omega \in \mathcal{V} \times \mathcal{V}$.

The *demand function* $\delta : \mathcal{V} \times \mathcal{V} \to \mathbb{R}_{\geq 0}$, which gives the amount of users associated with each node pair $(v_1^{\tau_1}, v_2^{\tau_2})$ in $\mathcal{G}$, is defined as $\delta(v_1^{\tau_1}, v_2^{\tau_2}) = \lambda(\{r \in R : d(r) = (v_1, v_2) \text{ and } t(r) = \tau_1 = \tau_2\})$. Note that this definition implies that $\delta(v_1^{\tau_1}, v_2^{\tau_2})$ can take nonzero values only when $\tau_1 = \tau_2$, i.e. when the two nodes are from the same isolated subgraph.

A *(path) flow* h on $\mathcal{G}$ is a vector in $\mathbb{R}^{|\mathcal{P}|}_{\geq 0}$ indexed by each path $p^\tau \in \mathcal{P}$ of $\mathcal{G}$. It is said to be *feasible* if for each node pair ω of $\mathcal{G}$, $\sum_{p^\tau \in \mathcal{P}_\omega} h_{p^\tau} = \delta(\omega)$, i.e., it satisfies the demands. Let $\mathcal{H}$ be the set of all feasible flows and note the convexity and compactness of $\mathcal{H}$. For each edge e^τ of $\mathcal{G}$ we shall denote edge flow by $h_{e^\tau} = \sum_{p^\tau \ni e^\tau} h_{p^\tau}$. For a path $p \in P$ of G, we let $h_p = \sum_{\tau \in \mathcal{T}} h_{p^\tau}$, i.e. the summation of flows on all copies of the same path. Similarly for an edge $e \in E$, we let $h_e = \sum_{\tau \in \mathcal{T}} h_{e^\tau}$.

The latency as a result of congestion differs slightly from the traditional selfish routing in the sense that it is *nonseparable.* The latency on an edge of $\mathcal{G}$ depends not only on the flow through this edge, but also on flows through all its copies with other labels. Formally, the edge latencies induced by flow h on $\mathcal{G}$ is given by $\ell_{e^\tau}(h) = l_e(h_e) = l_e(\sum_{\tau \in \mathcal{T}} h_{e^\tau})$ for each edge e^τ of $\mathcal{G}$. Note that copies of the same edge will have the same latency. Then assuming the additive model, for each path p^τ of $\mathcal{G}$ we have that the path latency induced by flow h is given by $\ell_{p^\tau}(h) = \sum_{e^\tau \in p^\tau} \ell_{e^\tau}(h)$.

We shall call the quadruple $(\mathcal{G}, \delta, \ell, \mathcal{T})$ a *discrete instance* with respect to instance (G, d, l, t).

Same as in the general model, the *disutilities* in a discrete instance is given by the summation of the latencies and the penalties. However we do not have to distinguish between different users anymore when talking about disutilities (we did not make this distinction when defining flows either), since the penalties are now encoded into the disutility function of graph $\mathcal{G}$. Formally, the disutility of using edge e^τ induced by flow h on $\mathcal{G}$ is given by $\mu_{e^\tau}(h) = \ell_{e^\tau}(h) + \tau_e$. Similarly path disutilities are given by $\mu_{p^\tau}(h) = \sum_{e^\tau \in p^\tau} \mu_{e^\tau} = \ell_{p^\tau}(h) + \tau_p$.

Further, we use $\mathcal{C}(h)$ to denote the *social cost* induced by flow h, that is $\mathcal{C}(h) = \sum_{p \in \mathcal{P}} \mu_p(h) h_p = \sum_{e \in \mathcal{E}} \mu_e(h) h_e$. Again we define the *auxiliary cost* for any flows h^1, h^2 as $\mathcal{C}_{h^1}(h^2) = \sum_{p\in\mathcal{P}} \mu_p(h^1)h^2_p = \sum_{e\in\mathcal{E}} \mu_e(h^1)h^2_e$.

In instance $(\mathcal{G}, \delta, \ell, \mathcal{T})$, a flow h^* is said to be at Wardrop equilibrium iff it is feasible and for all node pair ω of $\mathcal{G}$,

$$\left(h^*_p > 0 \implies \forall q \in \mathcal{P}_\omega,\ \mu_p(h^*) \leq \mu_q(h^*)\right), \quad \forall p \in \mathcal{P}_\omega. \tag{WP}$$

This is the so called Wardrop's principle [War52], and it can be formulated in terms of variational inequalities due to [Smi79]:

$$\mathcal{C}(h^*) - \mathcal{C}_{h^*}(h) \leq 0, \quad \forall h \in \mathcal{H}. \tag{WP}$$

Interrelationships. From the buildup of the discrete model, we already see its relationship to the general model. There are some very useful properties implied by this relationship which also explain our intuition in defining it this way. Now consider some finite instance (G, d, l, t) and its associated discrete instance $(\mathcal{G}, \delta, \ell, \mathcal{T})$. Observe the surjective mapping $g : F \to \mathcal{H}$ given by $g_{p^\tau}(f) = \lambda(\{r \in R : f(r) = p \text{ and } t(r) = \tau\})$ for each path p^τ of $\mathcal{G}$. In other words, $g_{p^\tau}(f)$ is the amount of users with penalty τ using path p according to flow f. Then the following properties can be verified trivially:

- for any flow $f \in F$, we have $f_e = g_e(f) = \sum_{\tau\in\mathcal{T}} g_{e^\tau}(f)$ for each edge e of G;
- for any flows $f^1, f^2 \in F$, we have $c_{f^1}(f^2) = \mathcal{C}_{g(f^1)}(g(f^2))$;
- $f \in F$ is at equilibrium iff $g(f)$ is at Wardrop equilibrium.

The reasoning behind setting up the discrete model lies in the fact that it possesses some nice characteristics which are of critical help in studying equilibria related properties of the general model.

Alphabet. Since we are dealing with general instances and discrete instances at the same time, it is not easy to make a distinction while being concise. One simple naming convention we adopt is that, we tend to use ordinary English letters for the general model and calligraphic and Greek letters for the discrete model.

Chapter 3

Properties of Equilibria

In this chapter, we will discuss equilibria related properties of our model defined in Section 2.3. First we will consider the existence of equilibria in Section 3.1. Then in Section 3.2 we give results for the uniqueness of equilibria. In Section 3.3 the price of anarchy in our model shall be discussed.

3.1 Existence of Equilibria

To study equilibria in our model, the most basic property in question is their existence. For finite instances, our discrete model is captured as a special case of the model studied in [AM81], due to the very general disutility function T allowed there. Thus the existence of equilibria is guaranteed by [AM81, Theorem 5.3]. However, their approach can not be generalized to infinite instances in our model. Instead, we observe that our general model fits into the very general definition of non-atomic games due to [Sch73]. In particular, [Sch73, Theorem 2] implies existence of equilibria in our model. With an attempt to make this thesis self-contained, we shall present a brief walk-through of their proof. However, we do not strive for preciseness nor completeness due to the depth of background involved, and because it is not a major topic of this thesis, nor is it our original work.

Non-Atomic Games. First we introduce the setup of non-atomic games and give the correspondence between non-atomic games and our general model. We have the set of players T represented by the real interval $[0, 1]$ endowed with Lebesgue measure λ. There is a finite activity pool of size n, that each user $t \in T$ can choose from. This corresponds to the set of paths in our model. We use the basis vector $e_i \in \mathbb{R}^n$, i.e. the vector in $\mathbb{R}^n$ with 1 on the ith dimension and 0 on all other dimensions, to denote an activity (a pure strategy of a player). Each player is confined to a nonempty subset of activities, denoted by $E(t) = \{e_{i_1}, \ldots, e_{i_k}\}$, as in our model where each user faces a set of paths connecting her origin to her destination. Let $P(t) = \text{conv}(E(t))$ be the set of available mixed strategies for player t, where conv stands for convex hull. Also let $P = \text{conv}(\{e_1, \ldots, e_n\})$ be the set of all possible mixed strategies, which is a superset of $P(t)$ for all t. A *T-strategy* is a measurable function $\hat{x} = (\hat{x}^1, \ldots, \hat{x}^n)$ from T to P. If a.e. $\hat{x}(t)$ is a basis vector, we say that $\hat{x}$ is a *pure* T-strategy. On the other hand, if a.e. $\hat{x}(t) \in P(t)$ we say $\hat{x}$ is feasible. Obviously a pure (and feasible) T-strategy corresponds to a (feasible) flow in our general model. Let $\hat{P}$ denote the set of all feasible T-strategies. Then the set $\hat{P}$ is compact and convex.

We need the payoff function to complete the construction of a game. We first introduce a utility function $\hat{u} : T \times \hat{P} \longrightarrow \mathbb{R}^n$. $\hat{u}^i(t_0, \hat{x})$ measures the utility of player t_0 choosing pure strategy e_i when every other player t chooses $\hat{x}(t)$. The utility function here plays the same role as the disutility function in our model which indicates the disutility of using a path for a user when a.e. user moves according to a flow.[1] So, the payoff of player t in T-strategy $\hat{x}$ is defined as $h_t(\hat{x}) = \hat{x}(t) \cdot \hat{u}(t, \hat{x})$.

There are two conditions to be satisfied by a non-atomic game in our discussion:

(a) For all players $t \in T$, $\hat{u}(t, \cdot)$ is continuous;

(b) For all feasible T-strategies $\hat{x} \in \hat{P}$ and $i, j = 1, \ldots, n$, the set $\{t \in T | \hat{u}^i(t, \hat{x}) > \hat{u}^j(t, \hat{x})\}$ is measurable.

Apparently the disutility function in our model meets both requirements if the latency function l is continuous.

Finally, a feasible T-strategy is said to be at *equilibrium* iff., a.e.,

$$\forall p \in P(t), \qquad h_t(\hat{x}) \geq p \cdot \hat{u}(t, \hat{x}),$$

which again corresponds to our equilibrium condition (EC) when $\hat{x}$ is a pure T-strategy.

Existence of Equilibria. By the setup of non-atomic games, it is easy to see that existence of pure T-strategies at equilibrium in non-atomic games will be a sufficient proof for existence of equilibria in our general model. In order to achieve this, the authors first proved existence of equilibria when allowing mixed strategies.

Theorem 3.1.1 ([Sch73, Theorem 1]). *A non-atomic game fulfilling conditions (a) and (b) admits a T-strategy at equilibrium.*

[1] The difference between *utility* and *disutility* can be simply eliminated with a negative sign.

Proof Sketch. For each player t and feasible T-strategy $\hat{x}$, define the set of best responses

$$B(t,\hat{x}) = \{p \in P(t) \mid \forall q \in P(t),\ p \cdot \hat{u}(t,\hat{x}) \geq q \cdot \hat{u}(t,\hat{x})\}.$$

Observe that $B(t,\hat{x})$ is convex and nonempty.

Claim 1. *For each player $t \in T$ the graph of $B(t,\cdot)$ is closed in $\hat{P} \times P$.*

For any t, consider sequences $\hat{x}_n \longrightarrow \hat{x}_0$ and $p_n \longrightarrow p_0$ that satisfy

$$\forall q \in P(t), \quad p_n \cdot \hat{u}(t,\hat{x}_n) \geq q \cdot \hat{u}(t,\hat{x}_n), \qquad n = 1, 2, \ldots$$

The continuity of $\hat{u}(t,\cdot)$ in (a) ensures that the inequality should hold in the limit, which proves Claim 1.

Then we define a set-valued function $\alpha : \hat{P} \longrightarrow \hat{P}$ as

$$\alpha(\hat{x}) = \{\hat{y} \in \hat{P} \mid \text{a.e. } \hat{y}(t) \in B(t,\hat{x})\}.$$

In other words, $\alpha(\hat{x})$ is the set of best mixed strategy profiles when the utility for each player is fixed at $\hat{u}(t,\hat{x})$. Note that any T-strategy $\hat{x}$ satisfying $\hat{x} \in \alpha(\hat{x})$ would also satisfy the equilibrium condition, and this is the direction where the proof goes.

Claim 2. *For each $\hat{x}$, $\alpha(\hat{x})$ is nonempty and convex.*

The convexity of $\alpha(\hat{x})$ follows from that of $B(t,\hat{x})$. As for nonemptiness, it is easy to construct an element of $\alpha(\hat{x})$. Define

$$T_i = \{t \in T \mid e_i \in E(t),\ \hat{u}^i(t,\hat{x}) \geq \hat{u}^j(t,\hat{x}), \forall j \in E(t)\},$$

i.e. T_i is the set of players for whom e_i is the best pure response. Apparently $\bigcup_{i=1}^{n} T_i = T$ and $\forall t \in T$, $e_i \in B(t, \hat{x})$. Also each T_i is measurable due to condition (b). Then we make them disjoint by letting $S_1 = T_1$ and $S_i = T_i \backslash (\bigcup_{j=1}^{i-1} T_j)$, $i = 2, \ldots, n$. Thus we have a T-strategy $\hat{y}$ in $\alpha(\hat{x})$ where $\hat{y}(t) = e_i$ for $t \in S_i$, $i = 1, \ldots, n$.

Claim 3. *The graph of α is closed in $\hat{P} \times \hat{P}$.*

This can be proved with the help of Claim 1 and [Aum65, Proposition 4.1]. Details are ignored here.

The properties of α in Claim 2 and Claim 3 fulfill the conditions of the Fan-Glicksberg fixed-point theorem, which guarantees the existence of $\hat{x}$ s.t. $\hat{x} \in \alpha(\hat{x})$. So the proof is finished. ∎

Now we look at the existence of pure equilibria. We need a further constraint on the utility function $\hat{u}(t, \cdot)$:

(c) For a.e. player $t \in T$, $\hat{u}(t, \hat{x})$ depends only on $\int_T \hat{x} = (\int_T \hat{x}^1(t)d\lambda, \ldots, \int_T \hat{x}^n(t)d\lambda)$.

Obviously the disutility function in our model also fulfills condition (c).

Theorem 3.1.2 ([Sch73, Theorem 2]). *If in addition to conditions (a) and (b), the utility function $\hat{u}(\cdot, \cdot)$ satisfies condition (c), then there exists a pure T-strategy at equilibrium.*

Proof Sketch. By Theorem 3.1.1, there is a T-strategy $\hat{x}$ at equilibrium. With condition (c), we know that a.e. $B(t, \hat{x}) = B(t, \hat{y})$ if $\int_T \hat{x} = \int_T \hat{y}$. Thus the direction of the proof is to show the existence of a pure T-strategy $\hat{y}$ such that $\int_T \hat{x} = \int_T \hat{y}$ and a.e. $\hat{y}(t) \in B(t, \hat{x})$. This is intuitively true, since we can reassign each player to one of her best pure responses, in a way that maintains $\int_T \hat{x}$. For a rigorous and detailed proof please refer to [Sch73] and [Aum65]. ■

Hence we have existence of equilibria in our general model as a direct application of Theorem 3.1.2.

Theorem 3.1.3. *Every instance (G, d, l, t) with continuous l admits a flow at equilibrium.*

One interesting fact to mention is that Theorem 3.1.2 essentially contains Nash theorem for finite games as a corollary [Sch73, Corollary]. This in some sense reflects the strength of this result.

3.2 Uniqueness of Equilibria

When a game admits more than one equilibria in general, which is obviously the case in our model, a natural question arising is how different those equilibria can possibly be and in which way they can differ. We will answer this question in this section. In fact we will show that for any instance (G, d, l, t), as long as the penalty function t is bounded, all equilibria of it will give rise to identical edge latencies. This implies that the disutility of each path will remain the same among all equilibria for each individual user. In other words, no user can really tell one equilibrium from another, assuming users can only perceive disutilities, since all equilibria look exactly the same to her. And this is what we mean by uniqueness of equilibria.

Similar studies have been conducted on traditional selfish routing, e.g. [Rou02, AM81]. We will extend their results to selfish routing with user preferences. First we will start with finite instances, i.e. when there are only finitely many distinct penalty assignments among all users. For finite instances, uniqueness of equilibria is a result of the fact that Wardrop equilibria in the discrete model can be formulated as solutions to a convex program. The approach is a standard way of showing existence as well as uniqueness of equilibria in a wide range of problems. In particular our proof for the following proposition is very similar to previous work on traditional selfish routing from [Rou02, Proposition 2.5.1].

Proposition 3.2.1. *For every finite instance (G, d, l, t) with continuous and nondecreasing l, if f^1, f^2 are flows at equilibrium, then $l_e(f_e^1) = l_e(f_e^2)$ for all edges e of G.*

Proof. Let us consider the corresponding discrete instance $(\mathcal{G}, \delta, \ell, \mathcal{T})$. From the interrelationships explored in the previous section, it is enough to show that any equilibrium flows h^1, h^2 of $(\mathcal{G}, \delta, \ell, \mathcal{T})$ must satisfy $l_e(h_e^1) = l_e(h_e^2)$ for each edge $e \in E$.

One important property of the disutility function in the discrete model is its integrability. This makes it possible to adopt techniques from convex optimization. Formally, define function $\mathcal{K} : \mathbb{R}^{|\mathcal{E}|} \to \mathbb{R}$ as

$$\mathcal{K}(h) = \sum_{e \in E} \left(\int_0^{h_e = \sum_{\tau \in T} h_{e^\tau}} l_e(x)dx + \sum_{\tau \in T} \tau_e h_{e^\tau} \right) \tag{KF}$$

where, with a slight abuse of symbols, we use h as an edge flow vector, that is, $(h_{e^\tau})_{e^\tau \in \mathcal{E}}$.

One can verify that the gradient of $\mathcal{K}$ at h is given by $(\partial\mathcal{K}/\partial h_{e^\tau}) = (\mu_{e^\tau}(h))_{e^\tau \in \mathcal{E}}$, i.e. the edge disutility vector due to flow h. Also observe the convexity of $\mathcal{K}$; moreover, each additive component inside the summation is convex. Hence the optimization problem

$$\text{minimize } \mathcal{K} \text{ over } \Gamma \cdot \mathcal{H} \tag{KP}$$

where Γ is the edge-path incidence matrix of $\mathcal{G}$ and $\Gamma \cdot \mathcal{H} = \{\Gamma h : h \in \mathcal{H}\}$, i.e. the set of feasible edge flows, is a convex program over a compact region, in which case h solves (KP) iff it satisfies (WP), i.e., it is at equilibrium for $(\mathcal{G}, \delta, \ell, \mathcal{T})$.

Thus, if h^1, h^2 are two equilibrium flows of $(\mathcal{G}, \delta, \ell, \mathcal{T})$, then both h^1, h^2, together with any convex combination $h^\theta = \theta h^1 + (1-\theta)h^2, \theta \in [0,1]$ will solve (KP). In other words, $\mathcal{K}(h^\theta) = \mathcal{K}(h^1) = \mathcal{K}(h^2)$. Hence $\mathcal{K}(h^\theta) = \theta\mathcal{K}(h^1) + (1-\theta)\mathcal{K}(h^2)$, i.e. $\mathcal{K}$ is linear between h^1 and h^2. Since $\mathcal{K}$ is the summation of convex components, each component must be linear between h^1 and h^2. This means that each l_e must take a constant value over $[h_e^1, h_e^2]$, thereby implying $l_e(h_e^1) = l_e(h_e^2)$ for each edge e of G. Therefore, because of the relationship between (G, d, l, t) and $(\mathcal{G}, \delta, \ell, \mathcal{T})$ given by the mapping g, we have proved the proposition. ■

The uniqueness of equilibria in infinite instances is not so straightforward however, as the growth of dimension to infinity will ruin the convex program formulation of equilibria. We tackle the infinity problem using a limiting approach. Intuitively one can construct a finite instance to *approximate* an infinite instance with certain precision. Then with the help of some convergence results, the uniqueness in infinite instances will become apparent. However, our uniqueness result for infinite instances will not depend on uniqueness in finite instances in Proposition 3.2.1, which may be somewhat unexpected. The proofs are just analogous in nature, and the only reason we include Proposition 3.2.1 is to help the reader better understand the following proofs.

So the first step is give a rigorous definition of *precision* of a finite instance approximating an infinite one.

Definition 3.2.2. For every instance (G, d, l, t), an instance (G, d, l, t') is an ϵ-approximation of (G, d, l, t) iff it is finite and a.e., $\forall e \in E, |t_e - t'_e| \leq \epsilon$.

Obviously, every instance (G, d, l, t) with bounded[2] t admits an ϵ-approximation for every $\epsilon > 0$. This can be achieved by partitioning the range of each t_e into fine enough intervals and rounding $t_e(r)$. In the following proposition, we shall explore further implications of this approximation.

Proposition 3.2.3. *For every instance (G, d, l, t), if (G, d, l, t') is an ϵ-approximation of it, then any equilibrium flow f^* of (G, d, l, t) will satisfy*

$$\textit{for any feasible flow } f, \quad c'(f^*) - c'_{f^*}(f) \leq 2(|V| - 1) \cdot \epsilon,$$

where $c'(.)$ and $c'_{(.)}(.)$ are the social cost function and auxiliary function of (G, d, l, t') respectively.

Proof. First of all, notice that modifications to penalty function t do not affect the set of feasible flows F of a general instance. Then for any feasible flows f^1, f^2 we have

$$\begin{aligned} |c_{f^1}(f^2) - c'_{f^1}(f^2)| &= \left| \int_R \left(u_{f^2(r)}(r, f^1) - u'_{f^2(r)}(r, f^1) \right) d\lambda \right| \\ &\leq \int_R \left| t_{f^2(r)}(r) - t'_{f^2(r)}(r) \right| d\lambda \\ &\leq \int_R \sum_{e \in f^2(r)} |t_e(r) - t'_e(r)| \, d\lambda \leq (|V| - 1) \cdot \epsilon. \end{aligned} \tag{3.1}$$

[2]Apparently unboundedness on subsets of zero measure does not matter according to Definition 3.2.2.

Using (3.1), we complete the proof for any feasible flow f by

$$c'(f^*) \overset{(3.1)}{\leq} c(f^*) + (|V|-1)\cdot\epsilon \overset{(EC)}{\leq} c_{f^*}(f) + (|V|-1)\cdot\epsilon \overset{(3.1)}{\leq} c'_{f^*}(f) + 2(|V|-1)\cdot\epsilon. \quad \blacksquare$$

Proposition 3.2.3 can be viewed as a relaxation of equilibrium condition (EC) or Wardrop's principle (WP). The factor $2(|V|-1)$ is not a concern since it will be a constant for a given instance and its approximations. Note that an ϵ-approximation is a finite instance by definition, and we have Proposition 3.2.1 as a result of Wardrop's principle applied on a finite instance. Now with a relaxed Wardrop's principle, we want to derive a relaxed yet strong enough result, which will imply the uniqueness of equilibria in general instances.

Proposition 3.2.4. *For every instance (G, d, l, t) with continuous and nondecreasing l and bounded t, if flow f^* is at equilibrium, then for any $\theta > 0$ there exists $\epsilon > 0$ such that any ϵ-approximation of (G, d, l, t), say (G, d, l, t'), satisfies $|l_e(f_e^*) - l_e(f'_e)| \leq \theta$ for all edges e of G, where f' is any equilibrium flow of (G, d, l, t').*

Proof. Fix an arbitrary equilibrium flow f^* of (G, d, l, t). For $\epsilon > 0$ to be determined later, let (G, d, l, t') be an ϵ-approximation of (G, d, l, t). We shall also consider the discrete instance $(\mathcal{G}, \delta, \ell, \mathcal{T})$ associated with (G, d, l, t').

From Proposition 3.2.3 we know for all feasible flows $f \in F$, we have $c'(f^*) - c'_{f^*}(f) \leq 2(|V|-1)\cdot\epsilon$. An equivalent way to express this in context of the discrete instance is

$$\text{for all } h \in \mathcal{H}, \quad \mathcal{C}(h^*) - \mathcal{C}_{h^*}(h) \leq 2(|V|-1)\cdot\epsilon \tag{3.2}$$

where $h^* = g(f^*)$ is flow f^* discretized for $(\mathcal{G}, \delta, \ell, \mathcal{T})$.

Let f' be an arbitrary equilibrium flow of (G, d, l, t'); hence $h' = g(f')$ is an equilibrium flow of $(\mathcal{G}, \delta, \ell, \mathcal{T})$. Now consider again the convex function $\mathcal{K}$ defined in (KF). Then h' minimizes $\mathcal{K}$ among all feasible flows of $(\mathcal{G}, \delta, \ell, \mathcal{T})$. Thus we have,

$$
\begin{aligned}
2(|V|-1)\cdot\epsilon &\overset{(3.2)}{\geq} -\sum_{e^\tau\in\mathcal{E}} \mu_{e^\tau}(h^*)(h'_{e^\tau} - h^*_{e^\tau}) = -\nabla\mathcal{K}(h^*)(h'-h^*) \\
&\geq \mathcal{K}(h') - \mathcal{K}(h^*) - \nabla\mathcal{K}(h^*)(h'-h^*) \\
&= \sum_{e\in E}\left(\int_0^{h'_e} l_e(x)dx - \int_0^{h^*_e} l_e(x)dx - \sum_{\tau\in\mathcal{T}} l_e(h^*_e)(h'_{e^\tau} - h^*_{e^\tau})\right) \\
&= \sum_{e\in E}\int_{h^*_e}^{h'_e} (l_e(x) - l_e(h^*_e))dx.
\end{aligned}
$$

To help make it clear, in the steps above we are actually bounding the gap between $\mathcal{K}$ and a tangent line. Note that no linear terms of $\mathcal{K}$ appear in the formulas as they vanished due to the second inequality. Since l_es are nondecreasing, each component inside the final summation is nonnegative. Therefore, for each edge e of G, we have

$$\int_{h^*_e}^{h'_e} (l_e(x) - l_e(h^*_e))dx \leq 2(|V|-1)\cdot\epsilon. \tag{3.3}$$

It would be beneficial to note the analogy between the proof technique here and that used in proving the uniqueness of equilibria in finite instances, as they share similar intuition.

Now it is time to determine an appropriate value for ϵ. Intuitively, one could already observe from (3.3) that for each edge e of G, $|l_e(h^*_e) - l_e(h'_e)| \to 0$ as $\epsilon \to 0$. To put it formally, for arbitrary $\theta > 0$, for each edge e of G, let ϵ^+_e, ϵ^-_e, and ϵ_e be defined as

$$\epsilon_e^+ = \begin{cases} \int_{h_e^*}^{y} (l_e(x) - l_e(h_e^*))dx & \text{if } \exists y,\ l_e(y) = l_e(h_e^*) + \theta, \\ 1 & \text{otherwise.} \end{cases}$$

$$\epsilon_e^- = \begin{cases} \int_{h_e^*}^{y} (l_e(x) - l_e(h_e^*))dx & \text{if } \exists y,\ l_e(y) = l_e(h_e^*) - \theta, \\ 1 & \text{otherwise.} \end{cases}$$

$$\epsilon_e = \frac{\min\{\epsilon_e^+, \epsilon_e^-\}}{2(|V| - 1)}.$$

Continuous and nondecreasing requirement of l_e guarantees $\epsilon_e > 0$. Thus by letting $\epsilon = \min_{e \in E}\{\epsilon_e\}$, inequality (3.3) implies $|l_e(h_e^*) - l_e(h_e')| \le \theta$ for each edge e of G, which essentially finishes the proof. ■

Conceivably, the uniqueness of equilibria in general instances follows as a trivial corollary from proposition 3.2.4, since for any pair of equilibrium flows f^1, f^2 of (G, d, l, t), it can be derived that $|l_e(f_e^1) - l_e(f_e^2)| \le 2\theta$ for any $\theta > 0$ on every edge e, which implies identity if we let $\theta \to 0$.

Theorem 3.2.5. *For every instance (G, d, l, t) with continuous and nondecreasing l and bounded t, if f^1, f^2 are flows at equilibrium, then $l_e(f_e^1) = l_e(f_e^2)$ for all edges e of G.*

3.3 Price of Anarchy

In this section, we are interested in the price of anarchy of general instances in our model, that is, the ratio of social cost of an equilibrium flow to that of an optimal flow. It characterizes the efficiency of equilibria in our considered model – in other words, the cost of missing a central coordinator. This work was motivated by similar studies in traditional selfish routing. In our model, the key generalization is the user-specific preferences or penalties, which influence both equilibria and social costs. We want to explore the roles of penalties in directing flows and incurring costs.

As is generally known, latency functions play a big part in the price of anarchy. We use the concept of β-function introduced in [CSSM04] to categorize and quantify the influence of latency functions.

Definition 3.3.1 ([CSSM04]). Let $\mathcal{L}$ be a family of continuous and nondecreasing latency functions. For every function $l \in \mathcal{L}$ and every value $v \geq 0$, define

$$\beta(v, l) = \frac{1}{vl(v)} \max_{0 \leq x \leq v} \{x(l(v) - l(x))\}.$$

In addition, define

$$\beta(l) = \sup_{v \geq 0} \beta(v, l) \quad \text{and} \quad \beta(\mathcal{L}) = \sup_{l \in \mathcal{L}} \beta(l).$$

In particular, the family of constant functions have β value 0, and the family of linear functions have β value 1/4. The definition of β-function implies that $l(x)y \leq \beta(x, l)l(x)x + l(y)y$, which will be used in proving the following theorem.

Theorem 3.3.2. *For every instance (G, d, l, t) with each l_e drawn from a continuous and nondecreasing function family $\mathcal{L}$, if f^* is a flow at equilibrium, then for all feasible flows f,*

$$\frac{c(f^*)}{c(f)} \leq \frac{1}{1 - \beta(\mathcal{L})}$$

where by convention we assume $\frac{0}{0} = 1$.

Proof. We start by separating the cost caused by latencies from that caused by penalties. For any two feasible flows f^1 and f^2, we have

$$\begin{aligned} c_{f^1}(f^2) &= \int_R (l_{f^2(r)}(f^1) + t_{f^2(r)}(r))d\lambda = \int_R l_{f^2(r)}(f^1)d\lambda + \int_R t_{f^2(r)}(r)d\lambda \\ &= \sum_{e\in E} l_e(f_e^1)f_e^2 + \int_R t_{f^2(r)}(r)d\lambda, \end{aligned} \tag{3.4}$$

and by using the definition of β-function, we have

$$\sum_{e\in E} l_e(f_e^1)f_e^2 \leq \beta(\mathcal{L})\sum_{e\in E} l_e(f_e^1)f_e^1 + \sum_{e\in E} l_e(f_e^2)f_e^2. \tag{3.5}$$

Thus pulling things together, we get

$$\begin{aligned} c(f^*) &\overset{(EC)}{\leq} c_{f^*}(f) \overset{(3.4)}{=} \sum_{e\in E} l_e(f_e^*)f_e + \int_R t_{f(r)}(r)d\lambda \\ &\overset{(3.5)}{\leq} \beta(\mathcal{L})\sum_{e\in E} l_e(f_e^*)f_e^* + \sum_{e\in E} l_e(f_e)f_e + \int_R t_{f(r)}(r)d\lambda. \end{aligned} \tag{3.6}$$

Since t is non-negative by assumption, and $\beta(\mathcal{L})$ is also non-negative by definition, we complete the proof by adding a non-negative term $\beta(\mathcal{L})\int_R t_{f^*(r)}(r)d\lambda$ to the right hand side of (3.6):

$$\begin{aligned} c(f^*) &\leq \beta(\mathcal{L})\sum_{e\in E} l_e(f_e^*)f_e^* + \beta(\mathcal{L})\int_R t_{f^*(r)}(r)d\lambda \\ &\quad + \sum_{e\in E} l_e(f_e)f_e + \int_R t_{f(r)}(r)d\lambda \\ &\overset{(3.4)}{=} \beta(\mathcal{L})c(f^*) + c(f). \end{aligned}$$

■

This bound on the price of anarchy is the same as that in traditional selfish routing [CSSM04], which reveals that introducing user-specific penalties does not degrade the efficiency of equilibria.

Chapter 4

Optimal Taxes

In this chapter, we study how helpful taxation can be in regulating equilibrium flows. In Section 4.1 we formalize the notion of edgewise tax and integrate it into our equilibrium model. Then we split our work on taxes into two parts: Section 4.2 will focus on taxes that induce equilibria with optimal social cost; Section 4.3 will discuss taxes that induce equilibria with optimal total latency.

4.1 Introducing Taxes

Knowing that pure selfishness might not induce flows that optimize public welfare, we are interested in edgewise taxation that helps drive selfish users into *desirable* flow patterns. Formally, for some instance (G, d, l, t), a *tax* b is a vector in $\mathbb{R}_{\geq 0}^{|E|}$ indexed by each edge $e \in E$, i.e. b_e presents the tax put on edge e by the administrator, just like the tolls on highways. Every user comes with a positive tax sensitivity, given by the *sensitivity function* $\alpha : R \longrightarrow \mathbb{R}_{>0}$ [1]. For each user $r \in R$ and each edge $e \in E$, the product $\alpha(r)b_e$ contributes to the disutility as an additive term. Thus tax sensitivity quantifies how a user balances between tax and latency in the ultimate formation of her disutility which in turn determines her choice of route. Recall that the penalty function t is so arbitrary that it allows us to incorporate tax and sensitivity into it. Therefore, in order to unify the representation, we define taxed penalty function $t^{\alpha b}$ as follows: for each user r, $t^{\alpha b}(r) = t(r) + \alpha(r)b$. Note that $t^{\alpha b}$ still qualifies for a penalty function, so we do not have to redefine equilibrium etc. for $(G, d, l, t^{\alpha b})$. Sometimes we shall call $(G, d, l, t^{\alpha b})$ a taxed instance.

In terms of tax sensitivity α, there are three cases in general: α is a constant function, in which case we can assume $\alpha = 1$ without loss of generality; α is not constant but takes on finitely many values; α takes on infinitely many values. The first case will be referred to as *homogeneous*, while the other two will be referred to as *heterogeneous*. In the homogeneous case, we shall ignore α and use notions like t^b directly.

Before moving on, we need make clear the exact meaning of *desirable* flow patterns. In this work, we shall discuss two different objectives to be optimized separately: the social cost, i.e. the total cost consisting of latencies and penalties, and the total latency.

[1] As usual we require that α is measurable.

4.2 Optimal Social Cost

We first investigate the possibility of using taxes to drive users into flow patterns that optimize the social cost. We begin with finite instances with homogeneous tax sensitivity. In this case, since user-specific penalties do not contribute to congestion in the system, intuitively the traditional marginal tax, where the tax of using an edge is equal to extra congestion caused to all users in the system, should serve for the purpose. Following the method in [BMW56], we verified that this is indeed the case, as can be seen from the following results.

Proposition 4.2.1. *For every finite instance (G, d, l, t) where l_e is differentiable and $xl_e(x)$ is convex for each edge e, a feasible flow f^o minimizes the social cost c iff f^o is at equilibrium for (G, d, l^*, t), where $l^*_e(x) = xl'_e(x) + l_e(x)$ and $l'_e(x) = \frac{d}{dx}l_e(x)$.*

Proof. Consider the discrete instance $(\mathcal{G}, \delta, \ell, \mathcal{T})$ associated with (G, d, l, t). Notice that feasible flow f^o minimizes the social cost $c(f) = \int_R u(r, f)d\lambda = \int_R (l_{f(r)}(f) + t_{f(r)}(r))d\lambda$ iff $h^o = g(f^o)$ minimizes $\mathcal{C}(h) = \sum_{p^\tau \in \mathcal{P}} \mu_{p^\tau}(h)h_{p^\tau} = \sum_{p^\tau \in \mathcal{P}} (\ell_{p^\tau}(h) + \tau_p)h_{p^\tau}$. Then we observe that the gradient of $\mathcal{C}$ is given by

$$\begin{aligned} \frac{\partial}{\partial h_{p^\tau}} \mathcal{C}(h) &= \sum_{e \in p} (h_e l'_e(h_e) + l_e(h_e)) + \sum_{e \in p} \tau_e \\ &= \sum_{e \in p} (l^*_e(h_e) + \tau_e) = \sum_{e \in p} (\ell^*_{e^\tau}(h) + \tau_e) = \mu^*_{p^\tau}(h). \end{aligned}$$

Note that the convexity of each $xl_e(x)$ implies the convexity of $\mathcal{C}$. Therefore flow h^o is a global minimizer of $\mathcal{C}$ iff,

$$\sum_{p^\tau \in \mathcal{P}} \mu^*_{p^\tau}(h^o)(h^o_{p^\tau} - h_{p^\tau}) \leq 0, \quad \forall h \in \mathcal{H},$$

which in turn is equivalent to Wardrop's principle for $(\mathcal{G}, \delta, \ell^*, \mathcal{T})$, i.e. the discrete instance of (G, d, l^*, t). This is also the criterion for f^o to be an equilibrium flow of (G, d, l^*, t). ■

With the observation from Proposition 4.2.1, we know that if we apply $l'_e(f^o_e)f^o_e$ as tax for each edge e, where f^o is a feasible flow of optimal social cost, then there exists at least one equilibrium flow, namely f^o, that minimizes the social cost, just as the following corollary concludes. On the other hand, note that such an optimal flow f^o always exists due to continuity of the social cost function $\mathcal{C}$ and compactness of the feasibility region $\mathcal{H}$.

Corollary 4.2.2. *For every finite instance (G, d, l, t) where l_e is differentiable and $xl_e(x)$ is convex for each edge e, let f^o be a feasible flow with optimal social cost. Then there exists a tax $b_e = l'_e(f^o_e)f^o_e$ such that (G, d, l, t^b) admits a equilibrium flow that minimizes the social cost for (G, d, l, t).*

Now recall our result on uniqueness of equilibria from Theorem 3.2.5. Then we exploit the uniqueness of equilibrium edge flows to strengthen Corollary 4.2.2.

Proposition 4.2.3. *For every taxed instance (G, d, l, t^b), if two equilibria f^1, f^2 share identical edge flows, then they give rise to the same social cost for (G, d, l, t).*

Proof. Since f^1 and f^2 give rise to identical edge flows, the disutilities will be equal between f^1 and f^2 for all available paths and hence the paths with the smallest disutility, for any user. Due to equilibrium condition (EC), a.e. user chooses the shortest path. Thus the social cost (with tax) for (G, d, l, t^b) induced by f^1 and f^2 are equal, i.e., $\int_R u^b(r, f^1) d\lambda = \int_R u^b(r, f^2) d\lambda$.

Then again because of uniqueness of edge flows, we have $\sum_e f_e^1 b_e = \sum_e f_e^2 b_e$. Together with the fact that $\int_R u^b(r,f)d\lambda = \int_R b_{f(r)}d\lambda + \int_R u(r,f)d\lambda = \sum_e f_e b_e + c(f)$, we have finished the proof. ■

Thus we naturally arrive at the following result on optimal social cost taxes in finite instances.

Corollary 4.2.4. *For every finite instance (G,d,l,t) where l_e is differentiable and strictly increasing, and $xl_e(x)$ is convex for each edge e, let f^o be a feasible flow with optimal social cost. Then there exists a tax $b_e = l_e'(f_e^o)f_e^o$ such that any equilibrium of (G,d,l,t^b) minimizes the social cost for (G,d,l,t).*

Remark. One might have noticed that the differentiability requirement of l_es in Corollary 4.2.4 can actually be dropped. By replacing derivatives with subderivatives and slightly modifying the proof, one could get the same result. We will not go into details of it in this thesis.

Next we turn to infinite instances. We shall build our results upon finite instances through another limiting argument. Same as for finite instances, we are working towards showing that any equilibrium flow of (G,d,l^*,t) essentially minimizes the social cost for the original instance. Again, we no longer have a convex formulation of the optimum.

We will explore further implications of the convergence result in Proposition 3.2.4. We know that if t is bounded and if each l_e^* is continuous and strictly increasing, then all equilibrium flows of (G,d,l^*,t) give rise to identical edge flows and there exists a sequence of finite instances converging to (G,d,l^*,t) in terms of equilibrium edge flows. The convergence of edge flows then results in the convergence of social cost. On the other hand, for this sequence of finite instances we know from Proposition 4.2.1 that their equilibrium flows give rise to smaller social costs than any arbitrary feasible flow. Conceivably, this relationship will hold in the limit, i.e. the limit of equilibrium social costs will remain smaller than (or equal to) the limit of the social costs induced by arbitrary feasible flows. This is equivalent to saying that the equilibrium of (G, d, l^*, t) induces the optimal social cost (with respect the original instance). The above argument is formally stated in Proposition 4.2.5 and Theorem 4.2.6.

Proposition 4.2.5. *For every instance (G, d, l, t) with bounded t, if for each edge e, l_e is continuously differentiable and l_e^* is strictly increasing, then for any $\theta > 0$ there exists $\epsilon > 0$ such that for any equilibrium flow f of (G, d, l^*, t) and any equilibrium flow $\bar{f}$ of any ϵ-approximation of (G, d, l^*, t), say $(G, d, l^*, \bar{t})$, we have $|c(f) - \bar{c}(\bar{f})| < \theta$, where $\bar{c}$ is the social cost function of $(G, d, l, \bar{t})$.*

Proof. For any $\theta > 0$, due to continuity of l, l^* and finiteness of the graph G, there exists $\theta_1 > 0$ s.t. if $|f_e - \bar{f}_e| < \theta_1$ for all edges e, then $|l_p^*(f) - l_p^*(\bar{f})| < \theta/3$ for all paths p and $|l_e'(f_e)(f_e)^2 - l_e'(\bar{f}_e)(\bar{f}_e)^2| < \theta/(3|E|)$ for all edges e.

Then from Proposition 3.2.4, we know there exists $\epsilon > 0$ s.t. for any ϵ-approximation of (G, d, l^*, t), say $(G, d, l^*, \bar{t})$, we have $|f_e - \bar{f}_e| < \theta_1$ for each edge e, where $\bar{f}$ is an arbitrary equilibrium flow of $(G, d, l^*, \bar{t})$. Without loss of generality we can assume $\epsilon < \theta/(3|V|)$. Thus we have, for a.e. user r,

$$\begin{aligned}
|u^*(r, f) - \bar{u}^*(r, \bar{f})| &= |\min_{p \in P_r}\{u_p^*(r, f)\} - \min_{p \in P_r}\{\bar{u}_p^*(r, \bar{f})\}| \\
&\leq \max_{p \in P_r}\{|u_p^*(r, f) - \bar{u}_p^*(r, \bar{f})|\} \qquad (4.1) \\
&\leq \max_{p \in P_r}\{|l_p^*(f) - l_p^*(\bar{f})| + |t_p(r) - \bar{t}_p(r)|\} < \frac{2}{3}\theta.
\end{aligned}$$

Hence

$$\begin{aligned}
|c(f) - \bar{c}(\bar{f})| &= |(\int_R u^*(r,f)d\lambda - \sum_{e\in E} l'_e(f_e)(f_e)^2) \\
&\quad - (\int_R \bar{u}^*(r,\bar{f})d\lambda - \sum_{e\in E} l'_e(\bar{f}_e)(\bar{f}_e)^2)| \\
&\leq \int_R |u^*(r,f) - \bar{u}^*(r,\bar{f})|d\lambda + \sum_{e\in E} |l'_e(f_e)(f_e)^2 - l'_e(\bar{f}_e)(\bar{f}_e)^2| \\
&\overset{(4.1)}{<} \theta.
\end{aligned}$$

■

Theorem 4.2.6. *For every instance (G, d, l, t) with bounded t, if for each edge e, l_e is continuously differentiable and $xl_e(x)$ is strictly convex, then any equilibrium flow f^o of (G, d, l^*, t) minimizes the social cost for (G, d, l, t).*

Proof. For the sake of contradiction, assume that there is a feasible flow f such that $c(f) < c(f^o) - \epsilon$ for some $\epsilon > 0$. Note that strict convexity of $xl_e(x)$ implies l^*_e is strictly increasing. Thus from Proposition 4.2.5 we can pick a finite instance $(G, d, l^*, \bar{t})$ such that for any equilibrium flow $\bar{f}$ of it, we have $|c(f^o) - \bar{c}(\bar{f})| < \epsilon/2$, where $\bar{c}$ is the social cost function of $(G, d, l, \bar{t})$. Without loss of generality we can assume a.e. $|t_e - \bar{t}_e| < \epsilon/(2|V|)$ and hence $|c(f) - \bar{c}(f)| < \epsilon/2$.

Putting everything together, we get $\bar{c}(f) < c(f) + \epsilon/2 < c(f^o) - \epsilon/2 < \bar{c}(\bar{f})$, contradicting Proposition 4.2.1 which says that $\bar{f}$ minimizes $\bar{c}$. ■

Remark. In fact Theorem 4.2.6 remains true when the strict convexity requirement on $xl_e(x)$ is weakened to just convexity, in consistency with Proposition 4.2.1. This can be achieved by elaborating Proposition 4.2.5, since we only need the fact that there exist one equilibrium (instead of all equilibria as in Proposition 4.2.5) for each instance in the convergent sequence, that converges in terms of the social cost.

With the above assumptions and propositions, we have enough to support the statement that the marginal tax drives users into flow patterns that optimize the social cost for general instances, as concluded by the following theorem.

Theorem 4.2.7. *For every instance (G, d, l, t) with bounded t, if for each edge e, l_e is continuously differentiable and $xl_e(x)$ is strictly convex, then there exists a tax $b_e = l'_e(f^o_e)f^o_e$ such that any equilibrium of (G, d, l, t^b) minimizes the social cost for (G, d, l, t), where f^o is an arbitrary equilibrium flow of (G, d, l^*, t).*

Proof. Direct result from Theorem 4.2.6, Theorem 3.2.5, and Proposition 4.2.3. One subtle point to notice is that strict convexity of $xl_e(x)$ implies that l is strictly increasing. ■

Seeing that edgewise taxes are so strong that they can in a sense overwhelm the variety of user-specific preferences, one might be tempted to generalize this result to heterogeneous tax sensitivities, as is the case in traditional selfish routings. However, this in general turns out to be beyond the reach of edgewise taxes. One simple counter example is enough to illustrate this limitation in the power of edgewise taxes. Consider an instance with two nodes and two parallel edges, with latency functions x and $1/2$ respectively. Half of the users (group A) have penalty 0 on both edges and have tax sensitivity 1; the other half of users (group B) have penalty $1/8$ on the linear latency edge and 0 on the other, and have tax sensitivity $1/4$. Figure 4.1 illustrates this situation, in which plus signs separate latencies and penalties. The path flow with optimal social cost can be formulated as the solution to the following

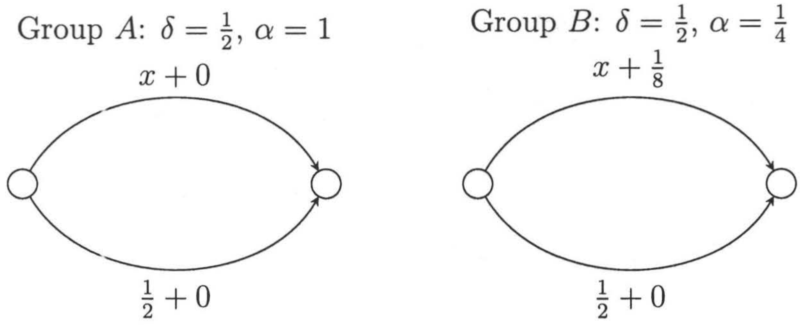

Figure 4.1: Example instance with heterogeneous tax sensitivities, for which no edgewise tax exists s.t. the induced equilibria have optimal social cost.

minimization problem:

$$\begin{aligned} \min x_{A1} \cdot (x_{A1} + x_{B1}) + x_{B1} \cdot (x_{A1} + x_{B1} + \frac{1}{8}) + (x_{A2} + x_{B2}) \cdot \frac{1}{2} \quad \text{s.t.} \\ x_{A1} + x_{A2} = \frac{1}{2} \\ x_{B1} + x_{B2} = \frac{1}{2} \\ x_{A1}, x_{A2}, x_{B1}, x_{B2} \geq 0, \end{aligned}$$

where x_{A1} means the amount of users in group A that travel through edge 1 (with linear latency function), and so on so forth. One can easily verify that the only solution to the above problem is $x_{A1} = x_{A2} = 1/4$, $x_{B1} = 0$, $x_{B2} = 1/2$, i.e. when half of the users in group A travel through the edge with linear latency, and all other users travel through the edge with constant latency. If we want to make this flow pattern to be at equilibrium using a set of edgewise tax, then we basically need to find b_1, b_2 such that

$$\begin{aligned} \frac{1}{4} + b_1 = \frac{1}{2} + b_2 \\ \frac{1}{4} + \frac{1}{8} + b_1 \cdot \frac{1}{4} \geq \frac{1}{2} + b_2 \cdot \frac{1}{4}, \end{aligned}$$

which obviously does not admit any solution. In other words, no edgewise taxation can drive users into optimal social cost flow in this small example.

4.3 Optimal Total Latency

In this section, we aim at a different target to optimize: the *total latency*. This objective function is more easily understandable, especially to readers familiar with traditional selfish routing, as it is not influenced by user penalties. Thus the criterion of a flow being optimal can be fully characterized by its edge flows, i.e. one only has to check its edge flows in order to tell whether a flow is of optimal total latency, contrary to checking the routing of every user.

With finite instances and finite tax sensitivities[2], i.e. when users can be grouped into a finite number of groups according to preferences and tax sensitivities, we notice the similarity between our model and traditional selfish routing with heterogeneous users. In particular, we observed that the proof techniques in [KK04a, Theorem 2] can be easily adapted to our scenario, which we shall present in the rest of this section.

Throughout this section we will be considering a finite instance (G, d, l, t) where l is *continuous* and *strictly increasing*. Also assume that tax sensitivity function α is *finite* and *strictly positive*. Let $(\mathcal{G}, \delta, \ell, \mathcal{T})$ be the associated discrete instance. Without loss of generality, we assume that all users with the same penalty $\tau \in \mathcal{T}$ have the same tax sensitivity which we denote by $\alpha(\tau)$, since otherwise we can further divide users by allowing $\mathcal{T}$ to be a multiset. Note that the symbol α here is used in a slightly different way than before, since originally it was a function of users. We

[2]As usual we say tax sensitivity α is finite iff it takes on finitely many values.

first introduce yet another formulation of equilibrium flows due to [AM81], which is equivalent to the Wardrop's principle we stated in (WP). They model traffic equilibria as solutions to a nonlinear complementarity problem [AM81, Proposition 4.1]. When adapted to our discrete model, it is as follows:

$$\begin{aligned}
(T_{p^\tau}(h) - y_\omega)h_{p^\tau} &= 0 && \forall \omega \in \mathcal{D},\ \forall p^\tau \in \mathcal{P}_\omega \\
T_{p^\tau}(h) - y_\omega &\geq 0 && \forall \omega \in \mathcal{D},\ \forall p^\tau \in \mathcal{P}_\omega \\
y_\omega\Big(\sum_{p^\tau \in \mathcal{P}_\omega} h_{p^\tau} - \delta_\omega\Big) &= 0 && \forall \omega \in \mathcal{D} \\
\sum_{p^\tau \in \mathcal{P}_\omega} h_{p^\tau} - \delta_\omega &\geq 0 && \forall \omega \in \mathcal{D} \\
h, y &\geq 0,
\end{aligned} \tag{CP}$$

where $\mathcal{D}$ is the set of all O-D pairs in $\mathcal{G}$. h and y are the variables in this problem, with h being the path flow vector, and y being the minimum disutility vector. With flow h and tax b, for each path $p^\tau \in \mathcal{P}$, the disutility T_{p^τ} is given as

$$T_{p^\tau}(h) = \ell_{p^\tau}(h) + \tau_p + \alpha(\tau)b_p = \sum_{e \in p}(l_e(h_e) + \tau_e) + \alpha(\tau)\sum_{e \in p} b_e$$

For convenience, let $\bar{\ell}_{e^\tau}(x) = \frac{l_e(x)}{\alpha(\tau)}$, $\bar{\tau}_e = \frac{\tau_e}{\alpha(\tau)}$. We alter the disutility function T_{p^τ} to be

$$T_{p^\tau}(h) = \sum_{e \in p}(\bar{\ell}_{e^\tau}(h_e) + \bar{\tau}_e) + \sum_{e \in p} b_e, \quad \forall p^\tau \in \mathcal{P}.$$

Notice that this modification does not affect feasibility of (CP). In particular the h solutions remain intact. Thus, after the modification of function T, the h solutions to (CP) still correspond to taxed equilibrium flows in $(\mathcal{G}, \delta, \ell, \mathcal{T})$ with tax b and sensitivity α.

Now we want to characterize an optimal total latency flow. As mentioned at the beginning of this section, the *optimality* can be fully described with edge flows. We assume that we are given an arbitrary feasible flow $h^o \in H$ of $(\mathcal{G}, \delta, \ell, \mathcal{T})$ which minimizes $\sum_{e \in E} h_e l_e(h_e)$ over the feasibility region H.[3] Then our aim is to find a tax b so that any solution (h, u) to (CP) will satisfy, $h_e = h_e^o$ for all edges $e \in E$. Hence the presence of tax b induces equilibrium flows that minimize the total latency. First observe the following important proposition.

Proposition 4.3.1. *If $h \in H$ is a feasible flow for $(\mathcal{G}, \delta, \ell, \mathcal{T})$, and h satisfies $h_e \leq h_e^o$ for all edges $e \in E$, then $h_e = h_e^o$, $\forall e \in E$.*

Proof. Assume for some edge $e \in E$, $0 \leq h_e < h_e^o$. Then since l_e is strictly increasing and nonnegative, we have $h_e l_e(h_e) < h_e^o l_e(h_e^o)$. As $h_e \leq h_e^o$, for all edges $e \in E$, we also know $h_e l_e(h_e) \leq h_e^o l_e(h_e^o)$, $\forall e \in E$. Thus we have $\sum_{e \in E} h_e l_e(h_e) < \sum_{e \in E} h_e^o l(h_e^o)$, a contradiction to the optimality of h^o. ■

We incorporate the constraints in Proposition 4.3.1 into (CP) to restrict the solutions, while still keeping the form of an nonlinear complementarity problem.

$$
\begin{aligned}
(T_{p^\tau}(h) - y_\omega) h_{p^\tau} &= 0 && \forall \omega \in \mathcal{D},\ \forall p^\tau \in \mathcal{P}_\omega \\
T_{p^\tau}(h) - y_\omega &\geq 0 && \forall \omega \in \mathcal{D},\ \forall p^\tau \in \mathcal{P}_\omega \\
y_\omega \Big(\sum_{p^\tau \in \mathcal{P}_\omega} h_{p^\tau} - \delta_\omega \Big) &= 0 && \forall \omega \in \mathcal{D} \\
\sum_{p^\tau \in \mathcal{P}_\omega} h_{p^\tau} - \delta_\omega &\geq 0 && \forall \omega \in \mathcal{D} \qquad \text{(BIG CP)} \\
b_e (h_e^o - h_e) &= 0 && \forall e \in E
\end{aligned}
$$

[3]Such an optimal flow exists due to compactness of H and continuity of l.

$$h_e^o - h_e \geq 0 \qquad\qquad \forall e \in E$$

$$h, y, b \geq 0.$$

Note that h_e is just a synonym for $\sum_{\tau \in \mathcal{T}} h_{e^\tau} = \sum_{\tau \in \mathcal{T}} \sum_{p \ni e} h_{p^\tau}$, as defined in the discrete model setup in Section 2.3. Also note that in (BIG CP) we now treat tax b as a variable of the problem. Then if (h^*, y^*, b^*) is a solution to (BIG CP), we know (h^*, y^*) is a solution to (CP) with

$$T_{p^\tau}(h) = \sum_{e \in p} (\bar{\ell}_{e^\tau}(h_e) + \bar{\tau}_e) + \sum_{e \in p} b_e^*, \quad \forall p^\tau \in \mathcal{P},$$

and from Proposition 4.3.1 we know $h_e^* = h_e^o$ for all edges $e \in E$. Therefore, h^* is a taxed equilibrium of $(\mathcal{G}, \delta, \ell, \mathcal{T})$ with tax b^*, and h^* minimizes the total latency. Moreover, since l is strictly increasing, due to the uniqueness of equilibria in Theorem 3.2.5, all equilibria give rise to identical edge flows. Hence all taxed equilibrium flows of $(\mathcal{G}, \delta, \ell, \mathcal{T})$ with tax b^* will minimize the total latency.

So our problem becomes clear: to prove the existence of a solution to (BIG CP), which will contain as a part of it an edgewise tax such that all taxed equilibria minimize the total latency. However, it is generally difficult to solve or to prove the existence of solutions for nonlinear complementarity problems. Luckily, the special structure of (BIG CP), namely Proposition 4.3.1, is capable of linearizing our problem. Since each function $\bar{\ell}_{e^\tau}(h_e)$ depends only on h_e and all solutions (h^*, y^*, b^*) of (BIG CP) will have $h_e^* = h_e^o$, $\forall e \in E$, replacing function $\bar{\ell}_{e^\tau}(h_e)$ with constant $\bar{\ell}_{e^\tau}(h_e^o)$ in each T_{p^τ} will not change the solution set of (BIG CP). Hence we have a linear complementarity problem equivalent to (BIG CP):

$$
\begin{aligned}
&(\sum_{e\in p}(\bar{\ell}_{e^\tau}(h_e^o)+\bar{\tau}_e)+\sum_{e\in p}b_e-y_\omega)h_{p^\tau}=0 && \forall\omega\in\mathcal{D},\ \forall p^\tau\in\mathcal{P}_\omega\\
&\sum_{e\in p}(\bar{\ell}_{e^\tau}(h_e^o)+\bar{\tau}_e)+\sum_{e\in p}b_e-y_\omega\geq 0 && \forall\omega\in\mathcal{D},\ \forall p^\tau\in\mathcal{P}_\omega\\
&y_\omega(\sum_{p^\tau\in\mathcal{P}_\omega}h_{p^\tau}-\delta_\omega)=0 && \forall\omega\in\mathcal{D}\\
&\sum_{p^\tau\in\mathcal{P}_\omega}h_{p^\tau}-\delta_\omega\geq 0 && \forall\omega\in\mathcal{D}\\
&b_e(h_e^o-h_e)=0 && \forall e\in E\\
&h_e^o-h_e\geq 0 && \forall e\in E\\
&h,y,b\geq 0.
\end{aligned}
\tag{BIG CP'}
$$

Now complementarity slackness conditions can be used to transform (BIG CP') into a primal-dual pair of linear programs. The existence of solutions follows henceforth.

Theorem 4.3.2. *(BIG CP') admits a solution.*

Proof. Due to complementarity slackness conditions, (BIG CP') is equivalent to the following primal-dual pair of linear programs:

$$
\min \sum_{p^\tau\in\mathcal{P}}(h_{p^\tau}\sum_{e\in p}(\bar{\ell}_{e^\tau}(h_e^o)+\bar{\tau}_e)) \quad \text{s.t.} \tag{LP}
$$

$$
\begin{aligned}
&\sum_{p^\tau\in\mathcal{P}_\omega}h_{p^\tau}\geq\delta_\omega && \forall\omega\in\mathcal{D}\\
&-h_e\geq -h_e^o && \forall e\in E\\
&h_{p^\tau}\geq 0 && \forall p^\tau\in\mathcal{P}
\end{aligned}
$$

and

$$\max \sum_{\omega \in \mathcal{D}} \delta_\omega y_\omega - \sum_{e \in E} h_e^o b_e \quad \text{s.t.} \tag{DP}$$

$$y_\omega - \sum_{e \in p} b_e \leq \sum_{e \in p} (\bar{\ell}_{e^\tau}(h_e^o) + \bar{\tau}_e) \qquad \forall \omega \in \mathcal{D},\ \forall p^\tau \in \mathcal{P}_\omega$$

$$y_\omega, b_e \geq 0 \qquad \forall \omega \in \mathcal{D},\ \forall e \in E$$

Since (LP) is feasible (e.g. h^o is a feasible solution), and its objective function is bounded from below by 0, the primal-dual pair (LP)-(DP) admits a solution (h^*, y^*, b^*), which is also a solution for (BIG CP'). ■

Hence we have proved the existence and showed the calculation method (by formulating and solving the primal-dual pair of linear programs) of a tax (which corresponds to the vector b in (DP)) that induces equilibrium flows of optimal total latency.

Theorem 4.3.3. *For every finite instance (G, d, l, t) where l is strictly increasing, if tax sensitivity α is finite and strictly positive, then there exists tax b s.t. any equilibrium of $(G, d, l, t^{\alpha b})$ minimizes total latency.*

For infinite instances, if one can show the existence of a convergent sequence of ϵ-approximations which admit a convergent sequence of optimal taxes, then it is easy to show that the converging point of taxes would be optimal for the original infinite instance. We strongly suspect it is true, but have been unable to verify it yet.

Chapter 5

Conclusions and Open Problems

In this work, we incorporated user preferences into selfish routing for the first time, and systematically studied several equilibria related properties. Though we have tried hard to study different aspects of our model, and to make our results general and hold under weak assumptions, there are still problems left unexplored, some of which have been briefly mentioned along the way. In this chapter, we shall review these open problems.

Boundedness of Penalty Function t

Recall that in the statement of many theorems and propositions in this thesis, there is an assumption that the penalty function t is bounded. This assumption first appeared in Section 3.2, in order to make a general instance "approximable". While this assumption is reasonable and acceptable in many real world problems we are trying to model, we must point out that it is not absolutely necessary in order to derive other results. We will explore this point in this section.

Recall our study on uniqueness of equilibria in Section 3.2. The main theorem, i.e. Theorem 3.2.5, was proved with a limiting argument. A crucial basis of the argument is the definition of ϵ-approximation in Definition 3.2.2 and its implication described in Proposition 3.2.3. One could easily see that Definition 3.2.2 is actually more restrictive than is needed by Proposition 3.2.3, and we shall discuss possible relaxations.

An immediate replacement for the definition of ϵ-approximation is the following:

Definition 5.1.4. For every instance (G, d, l, t), an instance (G, d, l, t') is an ϵ-approximation of (G, d, l, t) iff it is finite and for each edge $e \in E$, $\int_R |t_e(r) - t'_e(r)| d\lambda \leq \epsilon$.

Obviously Definition 5.1.4 is a less restrictive version of the ϵ-approximation in Definition 3.2.2, and it is also sufficient for the relaxed equilibrium condition in Proposition 3.2.3. In other words, this replacement of the definition does not weaken any other results in the thesis. Then the question is, what kind of penalty functions t admit ϵ-approximations according to the relaxed definition? Following the line of thoughts in bounded penalty functions, we generalize it to the concept of *subboundedness*, as is given in the following definition:

Definition 5.1.5. A measurable function $f : R \to \mathbb{R}_{\geq 0}$ is subbounded iff for any $\epsilon > 0$, there is a $k > 0$ such that $\int_{R_{f>k}} f(r) d\lambda \leq \epsilon$, where $R_{f>k} = \{r \in R : f(r) > k\}$.

To give a general feeling of what has actually been relaxed between boundedness and subboundedness, we give two simple example functions. First consider function $f^1 : R \rightarrow \mathbb{R}$ such that $f^1(r) = \frac{1}{r}$ when $r \neq 0$ and $f^1(0) = 0$. Apparently f^1 is neither bounded nor subbounded. Next we consider function $f^2 : R \rightarrow \mathbb{R}$ given by $f^2(1) = 0$, $f^2(r) = n$ when $r \in [S_n, S_{n+1})$, where $S_0 = 0$ and $S_n = \sum_{i=1}^{n} \frac{1}{2^n}$, $n = 1, 2, \ldots$. It is easy to verify that f^2 is not bounded (even when allowing zero-measure unboundedness) but it is subbounded.

Simply put, the subboundedness in Definition 5.1.5 requires that the unbounded portion of a function f is negligible in the sense of integration. Since we will only be looking locally at properties of penalty functions instead of considering the entire instance in the subsequent discussion, for simplicity, we sometimes use the term *ϵ-approximation* on two penalty functions instead of two problem instances, e.g. we may say t' is an ϵ-approximation of t. Next we explore the relationship between approximability, subboundedness, and Lebesgue integrability.

Proposition 5.1.6. *For a given penalty function t, the following three statements are equivalent:*

1. *For each e, t_e is subbounded;*
2. *t admits an ϵ-approximation for every $\epsilon > 0$;*
3. *For each e, $\int_R t_e(r)d\lambda$ is finite.*

Proof. **1 $\implies$ 2.** For any given $\epsilon > 0$, for each e, pick k_e such that $\int_{R_{t_e > k_e}} t_e(r)d\lambda \leq \epsilon/2$. Then define t'_e as

$$t'_e(r) = \begin{cases} \lfloor \frac{t_e(r)}{\epsilon/2} \rfloor \cdot \frac{\epsilon}{2} & t_e(r) \leq k_e, \\ 0 & \text{otherwise.} \end{cases}$$

Apparently t'_e is finite and $|t_e(r) - t'_e(r)| \leq \epsilon/2$ for any $r \in R \setminus R_{t_e > k_e}$. Thus we have

$$\begin{aligned}\int_R |t_e(r) - t'_e(r)| d\lambda &= \int_{R_{t_e>k_e}} |t_e(r) - t'_e(r)| d\lambda + \int_{R \setminus R_{t_e>k_e}} |t_e(r) - t'_e(r)| d\lambda \\ &= \int_{R_{t_e>k_e}} |t_e(r)| d\lambda + \int_{R \setminus R_{t_e>k_e}} |t_e(r) - t'_e(r)| d\lambda \\ &\leq \epsilon,\end{aligned}$$

i.e. t' is an ϵ-approximation of t.

2 $\implies$ 3. Let t' be an ϵ-approximation of t. From Definition 5.1.4 we have for each e

$$\int_R t_e(r) d\lambda - \int_R t'_e(r) d\lambda \leq \int_R |t_e(r) - t'_e(r)| d\lambda \leq \epsilon.$$

Because t' is finite by definition, it implies that $\int_R t'_e(r) d\lambda$ is finite for each e, so is $\int_R t_e(r) d\lambda$.

3 $\implies$ 1. Let t_e have a finite Lebesgue integral over R. For the sake of contradiction, let us assume t_e is not subbounded, i.e. there is an $\epsilon > 0$ such that for any $k > 0$, $\int_{R_{t_e>k}} t_e(r) d\lambda > \epsilon$. Consider the sequence of real-valued measurable functions $f_1, f_2, f_3, \ldots$ given by

$$f_n(r) = \begin{cases} t_e(r) & \text{if } r \in R_{t_e>n}, \\ 0 & \text{otherwise.} \end{cases} \qquad n = 1, 2, \ldots$$

Then obviously the pointwise limit of the sequence of functions $\{f_n\}$ as $n \to \infty$ is the zero function. Also, each f_n is dominated by t_e, that is for all $r \in R$, $|f_n(r)| \leq t_e(r)$, and by assumption t_e has a finite Lebesgue integral. Therefore, by Lebesgue's

Dominated Convergence Theorem, we have

$$0 = \int_R \lim_{n\to\infty} f_n(r)d\lambda = \lim_{n\to\infty}\int_R f_n(r)d\lambda = \lim_{n\to\infty}\int_{R_{t_e>n}} t_e(r)d\lambda.$$

However, by the assumption of being not subbounded we have $\int_{R_{t_e>n}} t_e(r)d\lambda > \epsilon > 0$ for all $n > 0$, which implies $\lim_{n\to\infty}\int_{R_{t_e>n}} t_e(r)d\lambda \geq \epsilon > 0$, a contradiction. ■

Proposition 5.1.6 accurately describes the required properties of penalty functions to be approximable. In other words, for all results in this thesis, we can replace each requirement of t being bounded by t being subbounded (or t having finite Lebesgue integral). And this is the best one can do to maintain the results without inventing a new proving approach.

However, there is still one question left unanswered: do we essentially need any assumptions on t at all, besides being measurable? What motivates this question is [Mil00, Proposition 3.3] which also gives a partial answer to it. An immediate adaptation of their result says, for networks with only parallel links and strictly increasing latency functions, as long as t is measurable, all equilibria give rise to identical edge flow. That is to say, subboundedness (or finite integrability) of the penalty function is not necessary to derive uniqueness of equilibria in these simple networks. We tried to verify the validity of their result in networks with general topologies, but unfortunately we have not been able to see it yet.

Efficiency of Equilibria

We have studied the price of anarchy in Section 3.3. There, we looked at the social cost, which combines latencies and penalties, as the objective to study. Our result says that the price of anarchy in traditional selfish routing remains valid after introducing user penalties. A very intuitive (and very unprecise) way of explaining this is that the newly introduced part can be viewed as a user-specific constant addition to latencies, which makes our model some sort of a "mixture" of constant latency networks and general latency networks. Because on constant latency networks the price of anarchy is one, which is the smallest possible by definition, the price of anarchy in our "mixed" model should not be anything greater than traditional selfish routing with general latency functions.

However, there are scenarios where system analysts are not concerned with users' satisfaction as a whole, which is represented by the social cost. Sometimes they are more interested in the network performance measured by total latency. Therefore, ratios such as

$$\frac{\sum_{e\in E} l_e(f_e^*)f_e^*}{\sum_{e\in E} l_e(f_e^o)f_e^o} \quad \text{and} \quad \frac{c(f^*)}{\sum_{e\in E} l_e(f_e^*)f_e^*} = \frac{\sum_{e\in E} l_e(f_e^*)f_e^* + \int_R t_{f^*(r)}(r)d\lambda}{\sum_{e\in E} l_e(f_e^*)f_e^*}$$

where f^* is a flow at equilibrium and f^o is an optimal total latency flow, might be of interest to some researchers. To further clarify things, the first ratio indicates how user preferences will affect and degrade the total latency as compared to the optimal case, and the second ratio indicates how overall satisfaction is related to the network performance at equilibrium. Studying these ratios is to some extent similar to the work in [CDR03, KK04b], except that we are facing a much more arbitrary "add-on" to latencies. How will the penalty function interact with these ratios? What restrictions are required on the penalty function to admit finite bounds on the ratios? We will leave these questions open to future studies.

Taxation in Infinite Instances

Back in Chapter 4, Section 4.3, we discussed taxes that drive users into optimal total latency flows. However, the whole discussion was based on finite instances, and we left the existence of optimal total latency taxes for infinite instances open. In this section we shall make one step closer to the possibility of extending the result to infinite instances, which we strongly suspect is feasible.

First recall our result in Section 4.3, particularly in Theorem 4.3.2. We know that for every finite instance (G, d, l, t)[1] and its associated discrete instance $(\mathcal{G}, \delta, \ell, \mathcal{T})$, with tax sensitivity function α, the optimal[2] tax b and the taxed equilibrium flow h can be formulated using the following (LP)-(DP) pair:

$$\min \sum_{p^\tau \in \mathcal{P}} (h_{p^\tau} \sum_{e \in p} (\bar{\ell}_{e^\tau}(h_e^o) + \bar{\tau}_e)) \quad \text{s.t.} \tag{LP}$$

$$\sum_{p^\tau \in \mathcal{P}_\omega} h_{p^\tau} \geq \delta_\omega \qquad \forall \omega \in \mathcal{D}$$

$$-h_e \geq -h_e^o \qquad \forall e \in E$$

$$h_{p^\tau} \geq 0 \qquad \forall p^\tau \in \mathcal{P}$$

[1]We will by default assume strictly increasing latency functions in this section.
[2]We shall use *optimal* to refer to *optimal total latency* in this section.

and

$$\max \sum_{\omega \in \mathcal{D}} \delta_\omega y_\omega - \sum_{e \in E} h_e^o b_e \quad \text{s.t.} \tag{DP}$$

$$y_\omega - \sum_{e \in p} b_e \le \sum_{e \in p} (\bar{\ell}_{e^\tau}(h_e^o) + \bar{\tau}_e) \qquad \forall \omega \in \mathcal{D},\ \forall p^\tau \in \mathcal{P}_\omega$$

$$y_\omega, b_e \ge 0 \qquad \forall \omega \in \mathcal{D},\ \forall e \in E$$

where h^o is a given optimal flow.

For every infinite instance (G, d, l, t) with t being bounded, there is a sequence of ϵ-approximations $\{(G, d, l, t^n)\}_{n=1,2,\ldots}$ converging to the original instance with $n \to \infty$ and $\epsilon \to 0$. For each one in the sequence, say (G, d, l, t^n), and its discrete instance $(\mathcal{G}^n, \delta^n, \ell^n, \mathcal{T}^n)$, the (LP)-(DP) pair would produce an optimal tax b^n. Then if somehow the sequence of taxes $\{b^n\}$ converges to some tax b as $n \to \infty$, one could easily show that b is an optimal tax for the original instance (G, d, l, t). This is formalized in the following proposition.

Proposition 5.3.7. *For every instance (G, d, l, t) with continuous and strictly increasing l and bounded t, let tax sensitivity α be finite and strictly positive. If there is a sequence of ϵ-approximations $\{(G, d, l, t^n)\}$ with $\epsilon \to 0$ when $n \to \infty$, such that it admits a sequence of optimal taxes $\{b^n\}$ converging to some tax b, then b is an optimal tax for (G, d, l, t).*

Proof. For each (G, d, l, t^n), assume it is an ϵ^n-approximation of the original instance. Let $\Delta b^n = \max_{e \in E}\{|b_e^n - b_e|\}$. Then due to our convergence assumption we have $\Delta b^n \to 0$ when $n \to \infty$. Obviously $(G, d, l, t^n + \alpha b^n)$ is an $(\epsilon^n + \Delta b^n)$-approximation of $(G, d, l, t + \alpha b)$, and $(\epsilon^n + \Delta b^n) \to 0$ when $n \to \infty$. Thus from Proposition 3.2.4

we know that any sequence of equilibrium flows of $\{(G, d, l, t^n + \alpha b^n)\}$ converges to the equilibrium flow of $(G, d, l, t + \alpha b)$ in terms of edge flows. By assumption b^n is an optimal tax for (G, d, l, t^n), for all $n = 1, 2, \ldots$, implying that their taxed equilibrium flows, i.e. equilibrium flows of $(G, d, l, t^n + \alpha b^n)$ for all n, share the unique edge flow, which is the optimal edge flow. Therefore their convergent point should be the same edge flow, which means b is an optimal tax for (G, d, l, t). ■

Thus, the problem is the existence of this sequence of ϵ-approximations. Let us look at (DP), the solution to which gives the optimal tax for a finite instance. As $\epsilon \to 0$ the number of constraints and the dimension of y are growing very fast, but the perturbation comes only from the RHS of the constraints, namely $\bar{\tau}_e$, and the magnitude of the perturbation is congruently converging to 0. It looks promising that the argument above implies the existence of a convergent sequence of b solutions. However, we are unable to see it at this moment, and will leave it open.

Bibliography

[AM81] H. Z. Aashtiani and T. L. Magnanti. Equilibria on a congested transportation network. *SIAM Journal on Algebraic and Discrete Methods*, 2(3):213–226, 1981.

[Aum65] Robert J. Aumann. Integrals of set-valued functions. *Journal of Mathematical Analysis and Applications*, 12:1–12, 1965.

[BMW56] M. Beckmann, C. B. McGuire, and C. B. Winsten. *Studies in the Economics of Transportation.* Yale University Press, 1956.

[CDR03] Richard Cole, Yevgeniy Dodis, and Tim Roughgarden. Pricing network edges for heterogeneous selfish users. In *STOC '03: Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 521–530, New York, NY, USA, 2003. ACM.

[CSSM04] José R. Correa, Andreas S. Schulz, and Nicolás E. Stier-Moses. Selfish routing in capacitated networks. *Mathematics of Operations Research*, 29(4):961–976, 2004.

[DS77] Carlos F. Daganzo and Yosef Sheffi. On stochastic models of traffic assignment. *Transportation Science*, 11(3):253–274, August 1977.

[FOV07] Babak Farzad, Neil Olver, and Adrian Vetta. Flow-free selfish routing. *submitted*, 2007.

[FT91] Drew Fudenberg and Jean Tirole. *Game Theory.* The MIT Press, 1991.

[KK04a] George Karakostas and Stavros G. Kolliopoulos. Edge pricing of multicommodity networks for heterogeneous selfish users. In *FOCS '04: Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS'04)*, pages 268–276, Washington, DC, USA, 2004. IEEE Computer Society.

[KK04b] George Karakostas and Stavros G. Kolliopoulos. The efficiency of optimal taxes. In Alejandro López-Ortiz and Angèle M. Hamel, editors, *CAAN*, volume 3405 of *Lecture Notes in Computer Science*, pages 3–12. Springer, 2004.

[KKVX07] George Karakostas, Taeyon Kim, Anastasios Viglas, and Hao Xia. Selfish routing with oblivious users. In Giuseppe Prencipe and Shmuel Zaks, editors, *SIROCCO*, volume 4474 of *Lecture Notes in Computer Science*, pages 318–327. Springer, 2007.

[Mil00] Igal Milchtaich. Generic uniqueness of equilibrium in large crowding games. *Mathematics of Operations Research*, 25(3):349–364, 2000.

[MMMT07] Marios Mavronicolas, Igal Milchtaich, Burkhard Monien, and Karsten Tiemann. Congestion games with player-specific constants. In Ludek Kucera and Antonín Kucera, editors, *MFCS*, volume 4708 of *Lecture Notes in Computer Science*, pages 633–644. Springer, 2007.

[MNS04] Patrice Marcotte, Sang Nguyen, and Alexandre Schoeb. A strategic flow model of traffic assignment in static capacitated networks. *Operations Research*, 52(2):191–212, 2004.

[Pap01] Christos Papadimitriou. Algorithms, games, and the internet. In *STOC '01: Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 749–753, New York, NY, USA, 2001. ACM.

[Per04] G. Perakis. The price of anarchy when costs are non-separable and asymmetric. In George L. Nemhauser and Daniel Bienstock, editors, *IPCO*, volume 3064 of *Lecture Notes in Computer Science*, pages 46–58. Springer, 2004.

[Pig20] Arthur Cecil Pigou. *The Economics of Welfare.* Macmillan, 1920.

[Rou02] Tim Roughgarden. *Selfish routing.* PhD thesis, Ithaca, NY, USA, 2002.

[Rou03] Tim Roughgarden. The price of anarchy is independent of the network topology. *Journal of Computer and System Sciences*, 67(2):341–364, 2003.

[RT02] Tim Roughgarden and Éva Tardos. How bad is selfish routing? *Journal of the ACM*, 49(2):236–259, 2002.

[Sch73] David Schmeidler. Equilibrium points of nonatomic games. *Journal of Statistical Physics*, 7(4):295–300, 1973.

[Smi79] M. J. Smith. The existence, uniqueness and stability of traffic equilibria. *Transportation Research Part B*, 13B:295–304, 1979.

[War52] J. G. Wardrop. Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers, Part II*, 1(36):352–362, 1952.

[YH04] Hai Yang and Hai-Jun Huang. The multi-class, multi-criteria traffic network equilibrium and systems optimum problem. *Transportation Research Part B*, 38(1):1–15, 2004.