UNIVERSITY OF CALIFORNIA,
IRVINE

Peer-to-peer and Collaborative Consumption of Supply in Transportation Systems

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Civil Engineering

by

Roger Lloret Batlle

Dissertation Committee:
Professor R. Jayakrishnan, Chair
Professor Amelia Regan
Professor Wenlong Jin
Professor Jean-Daniel Saphores

2017

DEDICATION

To my parents, Roger and Núria

Contents

List of Figures

List of Tables

ACKNOWLEDGMENTS

First, I would like to thank my family for letting me go so far away from home to pursue my research career.

Second, I acknowledge my advisor Prof. R. Jayakrishnan for having given me the freedom to do research on anything I wished, also for his street-smart advice on surviving in academia, and for financially supporting me during the last quarters of dissertation.

Third, I need to thank the Balsells Foundation for having provided me three years of full funding. Without this support, this dissertation would have never happened. Also, I have to acknowledge the Civil and Environmental Engineering Department and UC Connect for the financial support provided.

I also thank my ITS colleagues Neda, Robert, Mahdieh, Daisik, Jiangbo (Gabe), Felipe, Suman and Kate for all the advice and support received.

Also, Amine, Si-Yuan, Felipe and Vaibhav for doing such a great work during the Butterworth & Beall Enterpreneurship competition. It was a great experience!

I have to acknowledge Prof. Joseph Chow for his mentoring as a young faculty and also for providing me the opportunity to spend a great summer in NYU doing research and enjoying NYC together.

I would like to thank Prof. Ron Lavi for having welcomed me at the Technion. It was a great experience which gave a new perspective on how theoretical research is done.

I thank Prof. Kevin Roth for sharing with me his research results on value of urgency. I also thank Prof. Marc Fleurbaey and Prof. Benjamin Lubin for their insights in envy-free pricing and Dr. Steve G. Shelby for his advice on adaptive signal control.

I am indebted to Prof. Willfred Recker, Prof. Amelia Regan, Prof. Wenlong Jin, Prof. Jean-Daniel Saphores for the advice provided during the dissertation committee meetings.

Finally, I wish to thank Prof. Kenneth Small, Prof. Andreu Mas-Colell, Prof. Stergios Skaperdas, Prof. Michael McNally, Prof. Ivan Jeliazkov, Prof. Carolina Osorio, Prof. Igor Kopylov, Prof. Jean-Paul Carvalho, Prof. Michael Dillencourt, Dr. Yiheng Feng for their punctual advice received during the production of this dissertation.

CURRICULUM VITAE

Roger Lloret Batlle

EDUCATION

Doctor of Philosophy in Civil Engineering **2017**
University of California, Irvine *Irvine, California, USA*

BSc with MSc in Civil Engineering **2012**
Universitat Politècnica de Catalunya *Barcelona, Catalunya*

MSc in Transportation Engineering **2011**
Ecole Nationale des Ponts et Chausées *Champs-sur-Marne, France*

RESEARCH EXPERIENCE

Graduate Research Assistant **2012–2017**
University of California, Irvine *Irvine, California, USA*

TEACHING EXPERIENCE

Teaching Assistant **2016–2017**
University of California, Irvine *Irvine, California, USA*

REFEREED JOURNAL PUBLICATIONS

Envy-minimizing pareto efficient intersection control with brokered utility exchanges under user heterogeneity **2016**
Transportation Research Part B: Methodological

Envy-free Pricing for Efficient Adjustment of Service Order in Transportation Systems **2017**
The 22nd International Symposium on Transportation (ISTTT2017)

Using Bilateral Trading to Increase Ridership and User Permanence in Ridesharing Systems **2017**
Transportation Research Part E: Logistics and Transportation Review

P2P Ridesharing with Ride-back on HOV Lanes: Towards a Practical Alternative Mode for Daily Commuting **2017**
Transportation Research Record

Estimation of an inventory-theoretic model of mode choice in freight transport **2013**
Transportation Research Record

ABSTRACT OF THE DISSERTATION

Peer-to-peer and Collaborative Consumption of Supply in Transportation Systems

By

Roger Lloret Batlle

Doctor of Philosophy in Civil Engineering

University of California, Irvine, 2017

Professor R. Jayakrishnan, Chair

Transportation systems have been traditionally operated on a First-Come-First-Served (FCFS) fashion. FCFS consumption of supply occurs because it is accepted as a natural paradigm when the operators have no individual-specific information that allows consideration of any other serving order, and when users are assumed not to communicate among themselves. Thus, FCFS behaves as a status quo policy that is generally considered as fair, since it is presumed that all users are treated equally. We know though, that there exists heterogeneity in users' valuation of time and delay savings, and that the values may be different in different situations even for the same user. Taking advantage of smartphones and connected vehicle environments, it is now possible to include this user heterogeneity into operations in order to increase overall system efficiency and fairness, where efficiency refers to satisfaction of users. There are then possibilities of accomplishing this through exchanges among users with appropriate pricing, which can be determined by the users themselves to their satisfaction, so as to determine the order and extent of the utilization of supply. This new operational paradigm leads to collaborative consumption of supply.

This dissertation explores the idea of violating FCFS by allowing users to trade in real-time the part of supply that they effectively "own" while they are in a transportation

system. This de-facto ownership originates from the space-time region which each user rightfully controls, either due to their physical presence or due to reservations such as after purchasing a future trip from an operator. Attempting to answer the question of what pricing scheme would be fair and acceptable, leads this dissertation to introduce for the first time in transportation literature, the fundamental economic concept of envy-freeness. It can be taken as a pricing scheme as well as a user-behavior model. A resource allocation is said to be envy-free, when no agent feels any other agent's allocation to be better than their own, at the current price. An extension called dynamic envy-freeness is then developed for use in the domain of dynamic problems that the transportation field invariable pose, and a new family of envy-minimizing criteria are developed, namely the Constant Elasticity of Substitution Envy Intensity (CESEI) criteria, which strongly fits into the existing axiomatic body of Welfare Economics.

Several applications of collaborative consumption that breaks FCFS ordering are explored in this dissertation. First, the dissertation develops PEXIC, Priced EXchanges in Intersection Control, in which users can pay other users to reduce their waiting delays in a fair manner. This system is shown to be Pareto-efficient, envy minimizing and financially self-sustainable. Second, it studies new operational policies in highway control: parallel queue routing policies for bottleneck situations where the vehicles' lane-queue selections are the results of trades, and queue-jumping operations for exit lanes where vehicles can take forward spots in a queue by paying the overtaken vehicles in a fair fashion that achieves queue stability. Third, it proposes Peer-to-peer (P2P) ride exchange in ridesharing systems, in which trip property rights are transferred to users in such a way that they can trade their rides between each other. Finally, the dissertation models a P2P ridesharing system as a dual role market exchange economy, introducing a truthful pricing scheme which includes High-Occupancy-Vehicle (HOV) lane savings and uses a novel min-cost max flow formulation that guarantees users a ride-back, a complementarity in preferences never explored before.

The research does not attempt any elaborate examination of the social equity implications of such exchange-based systems with non-FCFS operations, but identifies some of such key issues and presents pointers for further study. It does not purport to take an advocacy position on transforming the transportation system operations to the newer paradigms, nor does it examine all the regulatory complications. The research does, however, demonstrate through modeling and analysis results from a variety of applications, that better system efficiency and user satisfaction can be achieved with the use of the proposed paradigms.

Chapter 1

Introduction

1.1 A new paradigm for transportation systems

This dissertation explores the idea of allowing users to trade in real-time the part of supply they "own" while they are in a transportation system. What is meant by "ownership" of transportation supply? Let's start with an example. Consider a vehicle driving along a highway: that vehicle and only that vehicle is occupying a location in space at a particular time, defining a trajectory in space-time. We can say that that vehicle is the de-facto "owner" of that portion of space at that time, since no other vehicle can occupy it. Perhaps as importantly, the vehicle may also be able to control a region in time and space, lawfully. An example would be a driver who drives slower than the average speed and delays a driver behind, effectively controlling certain regions on the road for certain time periods, purely because the driver is in front of those driving behind, which refers to the general First-Come First-Served (FCFS) nature of the consumption of transportation supply.

We know though, that there exists heterogeneity in users' preferences for travel time and delay savings. For instance, the vehicle behind may be in a hurry. Without delving into the practical aspects such as the requisite lane-changing for now, we can see that allowing the two vehicles to communicate and agree to swap positions, would lead to a more efficient outcome: the vehicle in a hurry would see its (highly valued) delay decrease, while the other one experiences extra delay, less valued than the former. Obviously, the second vehicle is worse off in this new situation, and therefore this operation must go together with a compensation, so as for both vehicles to reach an agreement to enact it. If enacted, it is akin to the traditional FCFS paradigm being broken. It may also require more than two users to agree before the operation can be carried out. Naturally, if the users are not more satisfied with a non-FCFS order, they would not agree to it, and it is rather intuitive to see that if it happens, it will be a more satisfying outcome for the agents, collectively.

The same concept extends to various types of transportation supply, of which users take de-facto possession of, and consume, where any user's control or ownership of any supply is mainly the result of the priority order of the user entering the system. This may refer to, for examples, a position in a queue at a signal, a reserved seat in a transit system, service order of a customer queue in a shared-ride vehicle dispatch system or at a pick-up point, etc. Obviously, any consideration of trading of supply that is 'owned' de-facto, is made possible only by newer technologies that facilitate an environment of rich information, communication and real-time computing. The necessary technologies currently exist or are expected to soon be available, as well. Such possibilities of trading of supply, however, bring up the necessity to properly re-examine the concept of supply in transportation, develop innovative methods for its use, and thus build a new paradigm on transportation system analysis and control.

Transportation supply is consumed instantly, since it cannot be stored. Moreover, it generally involves many interacting agents: multiple users or even operators and regulators, complicating the definition of ownership and the characteristics of potential trades. Arguably, real-time user decision-making can be limited due to cognitive capabilities or inadequacy to gauge the gains and losses on real-time, especially when the trading decisions happen very frequently. Moreover, for trading based on future costs, the spatio-temporal complexity of transportation systems necessitates predictive models for both the users and the operator. These predictive models allow the operator to monitor future states of the system. Once these future system states are known, the operator can communicate this information to users. As long as all parties rely on this information, users can evaluate the state of the system and choose among their trading alternatives.

Thus, a mediator or broker is implicitly needed for information and execution purposes. In the case where trades are just between users, this mediator will be the infrastructure manager. When the transportation supply is decided by one or more transportation operators, the mediator can be a public regulator. Supply is experienced by users in two separate cost components: non-monetary and monetary costs. Non-monetary costs are suffered while the user is being transported, although one could argue that, in case of purchasing the trip or transportation service, the user could foresee the non-monetary cost of the trip. In any case, this non-monetary cost is an internal cost which is non-transferable. Monetary costs, on the other hand are transferable and make the delay or comfort compensations possible.

Transportation supply can take two forms, which I will treat differently in this dissertation. When transportation services are booked in advance or these services are offered through modes with dedicated infrastructure or with limited uncertainty, the transportation supply "owned" by users can be treated as a private good. Examples are transferring the trip property rights to users in airplane flights, bus trips, or as proposed in this dissertation, P2P ridesharing systems. Road space, on the other hand, can be seen as a congestible (somehow rivalrous) public good. In economic parlance, this public good has non-excludability, because everyone can generally use it by either freely accessing system, like pedestrians or cyclists, or by paying a fixed annual lump sum tax like private vehicles, more tied to covering certain maintenance costs.

It is true, though, that public roads can be made "excludable" in a proactive manner. However, these policies focus on affecting the demand side. Some examples are toll pricing or cordon pricing like in Singapore or London. These solutions address congestion and environmental concerns on a very aggregate basis. They classify vehicles in categories and set up discriminating prices that depend on the volume of traffic and time of the day.

In this dissertation, the approach is different. Instead of enforcing excludability to control demand, I address the rivalry property. When considering the real-time individual level interaction, I claim that road space has the rivalry property, and this rivalry is well defined at the individual level: at every point of the space-time, only one vehicle can occupy that point. And more importantly, that ownership has a continuation in time and space, as the vehicle moves. This defines an FCFS supply consumption pattern since supply cannot be accumulated, is consumed instantly, and expires instantly. More importantly, this consumption pattern has always been considered a fair status-quo, since it was not possible to do differently, due to real-time information not being available. With the advent of smartphones and connected vehicle environments, this is not the case anymore. In that eventuality, the question is on whether the traditional FCFS pattern is indeed the most fair pattern for the society to use, or if more information on user-heterogeneity can be used to bring about usage patterns that are arguably fairer, and demonstrably more satisfying to the users. Attempting to provide initial answers to that question is the focus of this dissertation.

However, violating the FCFS rule leads to some agents having lower non-monetary costs and some others having higher of those costs, with regard to the status quo. This inequity implicitly needs some form of compensation between agents to address public acceptance and make the trading of supply acceptable, desirable, stable and fair. Users collectively agreeing to violate the natural FCFS consumption of transportation supply is what I call collaborative consumption. I define this collaborative consumption paradigm as the acquisition of space-time trajectories by users in exchange for compensation, which would normally be pecuniary. This idea is in line with the general definition of the term found in (Belk, 2014) in marketing literature.

The following figure summarizes Collaborative Consumption of Transportation Supply and conrasts it with Travel Demand Management:

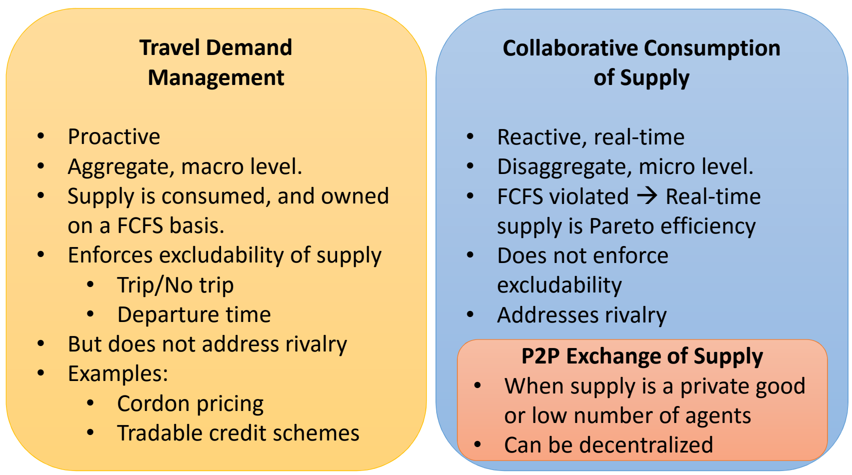


Figure 1.1: Travel Demand Management vs. Collaborative Consumption of Supply

With this in mind, our view of transportation supply can change radically. Implementing collaborative consumption, however, opens new challenges. Additional precautions need to be taken to prevent users manipulating the system for their personal gain. For instance, users' preferences are private information. The regulator is thus interested in having it revealed truthfully, to prevent economic arbitrage, inefficient future system states or inequities in utilities accrued by the users. This dissertation will address these regulations with mechanism design, a field of mathematical economics whose objective is to create mechanisms, or rules, to attain a particular economic objective, while inducing equilibria with desired properties.

Furthermore, newer concepts are also needed to model the behavior of the users. The traditional view of user behavior was in terms of the users' response to the collective transportation supply environment, and the choices offered. The variables that went into the models of the behavior pertained to one individual, and the collective environment faced by that individual. One individual's decision being influenced by another individual was rarely considered. Analyzing collaborative consumption, however, requires models of any user's consideration of other users on an individual basis. This dissertation proposes a method that is applicable for such cases, namely, the strive towards envy-free or minimum-envy outcomes from the user decisions. The concept of envy is based on the comparative valuation by a user, from his or her own standpoint, of the supply allocation to them and to another user. While the basic concept has existed in economics, it has not been used in the transportation field, and a significant extension of the concept to the dynamic case is necessary for modeling the user behavior under exchange-based supply consumption in transportation. This dissertation makes headway on both counts and demonstrates the use of the concept of envy in practical situations to model user behavior in collaborative supply consumption in transportation.

In summary, my research proposes to analyze the above-mentioned new paradigms in our understating of transportation supply and its consumption. First, it provides new fundamental concepts and frameworks for transportation systems analysis. Second, it studies practical applications where P2P exchange and/or collaborative consumption and quantify its effects. This dissertation is truly multidisciplinary and it intersects with transportation engineering, economics, operations research and computer science.

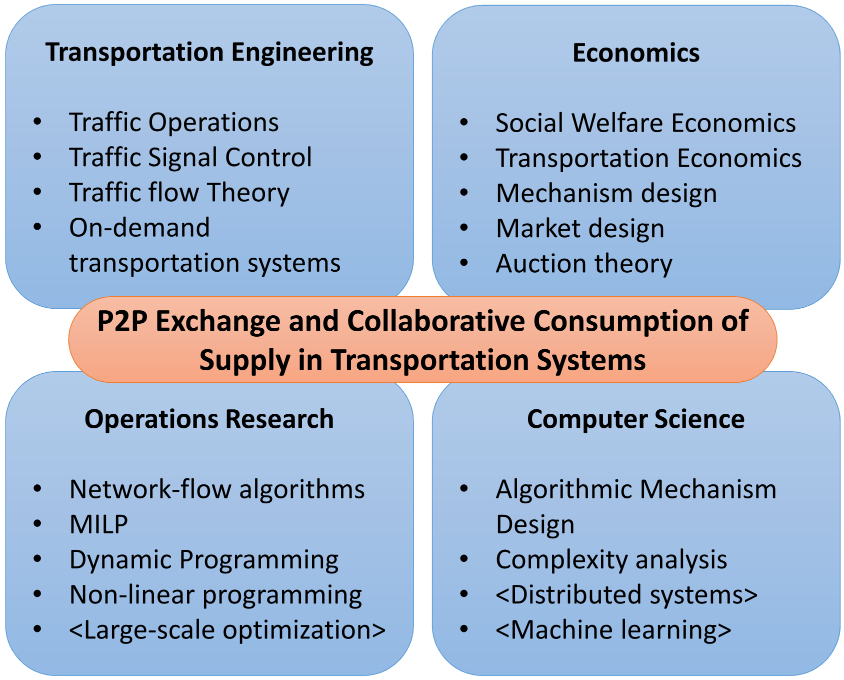


Figure 1.2: This dissertation intersects with four different disciplines

1.2 Dissertation Outline

The dissertation is organized in the following chapters:

- Chapter 2 provides the conceptual underpinnings on which the rest of the dissertation is supported. The idea and consequences of violating FCFS are further explored. A new supply-demand equilibration framework which includes P2P and collaborative consumption of supply is developed.
- Chapter 3 consists of a review on mechanism design and social welfare concepts used in the dissertation. It also includes contributions to social welfare literature such as Dynamic Envy-Freeness and a new criteria on envy minimization: the Constant Elasticity of Substitution Envy Intensity (CESEI) criteria family. This chapter

is based on Lloret-Batlle and Jayakrishnan (2017).

- Chapter 4 presents PEXIC, Priced EXchanges in Intersection Control. PEXIC is the first collaborative consumption traffic signal control algorithm. It considers vehicle heterogeneity in Value of Delay Savings. It allows high value vehicles to obtain lower average delays by compensating lower value vehicles which wait longer with a payment. This chapter is based on Lloret-Batlle and Jayakrishnan (2016). This dissertation presents an additional section not present in the published article, where the minimum-envy prices are shown to be approximately Bayesian Incentive Compatible.

- Chapter 5 explores collaboration consumption of supply in traffic operations. First, I propose collaborative dynamic queue routing for freeways and multiserver queue facilities. Under this policy, vehicles can collaboratively change lanes (or switch queues) such that the total cost of the platoon is minimized, in a fair and coalitionally stable manner. Second, a new type of traffic operations: queue-jumping operations for freeway exit ramps. Arriving vehicles are allowed to skip positions in a moving queue by paying the overtaken queued vehicles and other arriving vehicles in a way that the new queue order is Pareto efficient, fair (envy-free) and financially self-sustainable. The first section of this chapter is based on Lloret-Batlle and Jayakrishnan (2018), the second section is based on Lloret-Batlle and Jayakrishnan (2017).

- Chapter 6 studies applications of collaborative consumption in P2P ridesharing systems. I study the positive effects of transferring property rights of trips to users in a dynamic P2P ridesharing system. Users who join the system and do not find any ride available are allowed to buy a trip from users who arrived earlier and had not yet been picked up. This chapter is based on Masoud, Lloret-Batlle and Jayakrishnan (2017).

- Chapter 7 presents a parametric study on P2P dual-role exchange markets. These markets have the particular characteristic that agents can take either the role of consumer or a supplier. This is the case of P2P ridesharing systems, in which travelers can choose to be either drivers or riders. In this chapter, P2P ridesharing is conceived as an alternative mode commuting solo. This chapter is based on Lloret-Batlle et al. (2017).
- Chapter 8 provides the conclusions and further research.

1.3 What this dissertation does and what it does not

Researchers often have a preconceived idea of any new topic they tackle. This is often true when the new subject is of a multidisciplinary nature, where problems of conflicting semantics add up. By the same token, the readers may also develop notions about a research thesis which may result in confusing thoughts. This section provides a summary of this dissertation's goals, as well as goals that were not pursued. The following paragraphs are motivated by many of the questions received during the dissertation research years, in many academic conferences as well as at seminars I have done on and off-campus.

This section is placed early in the text, for the sake of those readers who may have already developed questions about the ramifications of the dissertation based on the introduction itself, so as to answer some concerns they may have, and help them in seeing the rest of the dissertation in the proper light. The readers are requested to return to this section as and when needed while going through the chapters, especially if the tone of discussions and the practical details of the presented applications appear to be dissimilar, and if the overall logic behind the research progress is rendered unclear by that. It is fair to say that this was mostly by design, and was caused by the nature of the subject that has implications in various directions, some that are studied here and some that are not.

The main goal of this dissertation is to explore a new paradigm in the operation of transportation systems, which we call Peer-to-peer and Collaborative Consumption of Transportation Supply. This consists mainly of exploring how this paradigm outperforms FCFS operation, by providing a benchmark for each of a few applications. The study has been carried out on five applications, with the purpose of displaying a broad scope of the matter, rather than of concentrating on just one or two applications and missing out on important contextual colorations of the paradigm.

The general goal is addressed in the following manner: the system is modeled as an FCFS system, where no real-time value of information (i.e. VOT) is requested from users. Then, the system is first modeled with the designed collaborative consumption mechanism or control policy. For both systems, the same metrics are evaluated and compared. In some cases, like in Chapter 4, new metrics had to be developed. These metrics concern relevant designer objectives such as social welfare, fairness, revenue or coalitional stability.

For each of the applications, an ad-hoc collaborative consumption mechanism has been designed. When designing an economic mechanism, the designer needs to choose a set of goals to be satisfied, or at least approximated. More often than not, there are impossibilities and trade-offs between these goals. One example is the trade-off between efficiency and revenue found in Chapter 7. The goal selection criteria has an application-specific component, but also attempts to broaden the scope of the dissertation. In Chapter 4 the focus was on fairness, which also turned out to be approximately incentive-compatible while in section 5.1, I chose to explore coalitional stability, instead of incentive compatibility, as the goal.

Concerning the operational scope of this dissertation, with the exception of the ridesharing chapters 6 and 7, the focus was on analyzing local traffic phenomena. This includes one traffic intersection (Chapter 4), one freeway exit ramp and one multiserver queue or freeway link (Chapter 5). Recognizing the increased complexity of the corresponding extension to network-level applications, the focus was instead directed towards analyzing different physical systems, also as an excuse to study different goals (efficiency, incentive compatibility, fairness, revenue, coalitional stability).

The benchmarks presented in this dissertation differ on the modeling details and on the implementation depth. Chapters 4 and 6 incorporate detailed models which are close to being implementation-ready, and state-of-the-art as far as the underlying algorithms are concerned. For the rest of chapters, the model complexity has been reduced in order to explore and better understand new applications from a normative point of view. This is especially the case of 5.

The contents of chapter 4 have been by far the most polemic. The main goal of this chapter was to compare social welfare and fairness of an intersection operated with delay minimization, against the same intersection operated with collaborative consumption of supply. The conclusions are that under this new control scheme the intersection is operated more efficiently and in a more fair manner. Eventual public acceptance of the control system is addressed by showing that the new control Pareto improves upon delay minimization control (in efficiency), and fairness is improved regardless of the valuation of delay savings of the user. This chapter does not address strategic participation of agents such as evolutionary voluntary participation, but it does address the prevention of economic arbitrage. In other words, it assumes full participation in the new scheme. Finally, the minimum-envy prices are found to be approximately Bayesian Incentive Compatible.

Table 1.1 summarizes the goals achieved for each application:

Application	Eff.	Rev.	Fairness	IC	Stability	IR
Traffic Signal Control (Ch. 4)	X		X	X		
Cooperative Queue routing (Sec. 5.1)	X				X	X
Queue jumping operations (Sec. 5.2)	X		X			X
P2P Exchange in Dynamic Ridesharing (Ch. 6)	X			X		X
Day-to-day Ridesharing (Ch. 7)		X		X		X

Table 1.1: Economic goals per application. (IC: Incentive Compatibility, IR: Individual Rationality)

Another out-of-scope concern raised during conferences are income inequalities. Income is disregarded in two dimensions, both explained next. First, all models presented in this dissertation have quasi-linear utility functions: they do not incorporate income effects, since it is considered that those payments are small with respect to the agents' incomes. This is especially true due to the operational scope of this dissertation, which is constrained to local transportation phenomena or local rides, where delay savings are generally small and so, as well, are the corresponding payments.

Second, the models are budget-free, in that they do not incorporate a system-level budget constraint either. Making the models budget-free eventually gives room to different monetary system implementations. For instance, the monetary units could come from either users' income or could also come from a close credit system, e.g. tradable credit scheme, where users' income does not play any role. Budget can then be controlled on a different, independent layer for a network-level model, aiming a hierarchical implementation, so common in traffic control.

As should be evident, this dissertation does not set out to explore every aspect of the paradigms presented herein. Though mentioned in various places, even preliminary, let alone complete, exploration of certain issues such as related to social justice, and the regulatory framework needed to preserve it, are out of the scope of this work. Nevertheless, the chapters do demonstrate that the intuitive ideas one may develop of injustice or unfairness are not necessarily valid, and even that they are mostly invalid, in much of the contexts studied.

Having said that, this dissertation is still not meant to help crystallize a position of advocacy for the paradigms explored. Quite the contrary. It merely shows what can happen if transportation systems were to operate under paradigms different than an FCFS-based supply consumption pattern. Furthermore, it provides a basis to view the system with associated gains and losses if any bottom-up entrepreneurs or system-control agencies explore such paradigms, purely based on availability of data and ease of communication technology use. The fact remains that if there is inefficiency to exploit in the system, and thereby make money, eventually it will get exploited. For this reason, there are indications that changes along similar lines as discussed in this dissertation may have already started happening.

Rather than advocate such changes, the dissertation remains as one that both envisages possibilities and studies their impacts, the intent being for us to be insightfully prepared for tackling developments in such directions. It is not for the engineer to predict social and systemic changes. Analyzing possibilities of change for their effects, is indeed in the engineer's domain, nonetheless. That is the focus of the work here. It is left for the society with its political representatives, to effect, or even control, fundamental changes in transportation systems operations. The hope is that this dissertation provides input for the society to take any stand for or against such systemic transformation, or to channel the changes in appropriate directions, with proper concepts and results to back up the judgments on when such new paradigms are beneficial, when not, and to what degree in either case.

Chapter 2

Conceptual Underpinnings

2.1 Introductory Example

This chapter first develops an introductory example to give a practical context to the new supply-consumption paradigms presented in the first chapter, to facilitate a discussion of the conceptual underpinnings, before proceeding to more elaborate application contexts in the latter chapters.

Consider that there is a one lane, two-way highway with two vehicles: 1 and 2. Vehicle 1 is in front of vehicle 2 since it departed earlier. Vehicle 1 drives slower than vehicle 2, which drives at maximum allowed speed, is in a hurry and would like to overtake vehicle 1. Vehicles are approaching a section where overtaking is forbidden. However, vehicle 2 cannot overtake since the maneuver would finish in the forbidden section, which is dangerous in case someone else comes from the other direction. Under FCFS operations, the order would be maintained and vehicle 2 would suffer an externality cost. However, if vehicle 1 were to slow down, vehicle 2 could overtake it safely and spend less time on the road or arrive on time. Obviously, this may require some sort of compensation to vehicle

1, that makes it worthwhile for that vehicle to cooperate in this maneuver.

The above operation requires five important points to be addressed:

1. Assessment of communication between agents,
2. Appropriate measuring of the externalities,
3. Alternative choice set definition,
4. Valuation of the externalities (and choices) from the agents, and
5. Definition of a valid mechanism.

The first point is crucial: if there is no communication, no trading is possible. Moreover, this communication protocol needs to be responsive enough and have enough range to cover the agents that are of concern in such trading. From a financial point of view, the communication system must guarantee small transaction costs. If transaction costs were high, the efficiency gains from the trade would be wasted, or even worse, no trade at all could ever happen. I assume in this dissertation that agents possess such communication protocol. Some valid examples, application-specific as they are, would be smartphones and connected vehicle environments. For complex interactions, in which many agents are involved, such as a traffic signal, communication with a central operator, facilitator or broker may be necessary. For simpler interactions, a common, standardized pure-P2P-communication protocol based on connected vehicles would be enough.

The second aspect is a classic issue in microeconomics. If externalities cannot be really measured, there is no reason for agents to trade, since they cannot agree on what to trade in. However, if good communication between vehicles and/or infrastructure is satisfied, it is possible to estimate their externalities in delay or comfort. Again, the externalities may be calculated by a controller or central operator, or in a decentralized fashion if vehicle interactions are sparse, like in the example above. Once the system has a way to estimate the externalities, the set of different trading alternatives or users' choices users can be determined, as in the third point above. These alternatives will mainly be defined by time and space constraints.

The consideration of user heterogeneity being the main motive behind proposing collaborative consumption of supply, the fourth point above requires special attention. Externalities are valued individually by every user, and this valuation per unit of externality is generally a private parameter which has to be elicited from users. How to address this issue, and the different methods used are explained in the next sections and chapters. This is both a behavioral issue from the standpoint of the users and a system-objective issue from the operators' or society's standpoint. It must be noted, that the use of the word externality also includes pecuniary externalities such as when private goods are traded.

Finally, a valid trading mechanism has to be designed. This mechanism should attempt to guarantee a better future state of the system. In this dissertation, a better state is defined as one with an increase in efficiency, i.e. utilitarian social welfare and increase in fairness. For this, reduction of envy described in Chapter 3 offers a useful vehicle. These concepts will be explored next in the following sections based on the above example context.

Note also, that the mechanism should ensure correct elicitation of private information and make sure that users do not benefit from the system at the expense of other users. An example is economic arbitrage, such as when trips are induced for primary consumption and not as derived demand from another economic activity. A simple case would be drivers getting on a highway purely to make money from payments for not slowing down others, an eventuality to be avoided. This underscores the need for proper elicitation of private information.

Let's look again at the introductory example given above. Assuming that the communication protocol does not limit the possibilities of trade, the first step is to calculate the magnitude of the externalities to build next the set of alternative options. In the example, the only meaningful action is vehicle 1 decelerating to allow vehicle 2 overtake it. From many levels of deceleration possible, only the minimum one such that vehicle 2 can accelerate is optimal. The rest would add extra inefficiencies due to slowing down vehicle 1 excessively. Thus, there are two alternatives: first, not changing the order arising from FCFS, and second, optimally decelerating vehicle 2 and letting vehicle 1 overtake it. Taking as the reference point the FCFS allocation in which both agents' payoff is zero, the delay to vehicle 1 increases by d_1 and the delay to vehicle 2 decreases by d_2. d_2 is considered greater than d_1, as seen in Figure 2.1.

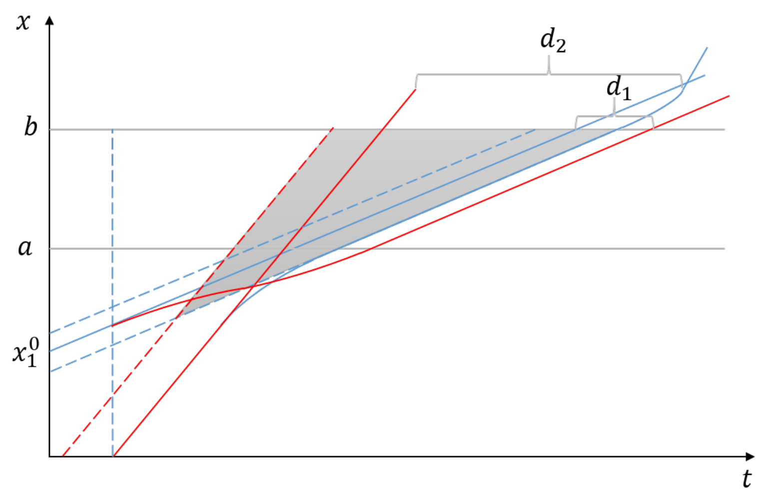


Figure 2.1: Introductory example: space-time ownership, trading alternatives, externalities

The next step is to know how much each agent values his or her potential allocation. Let us assume that the delay cost and prices are valued separately and additively. This is an assumption of quasilinearity and it is common in mechanism design. More details on this assumption will be given in the next chapter. Furthermore, let us assume that each agent values every unit of delay θ_i (\$/h) in a linear fashion. The utility specification is then called linear. Since vehicle 2 is in a hurry, $\theta_2 > \theta_1$.

Several trading mechanisms for this setting will be examined next. The agents will have to decide if they optimally switch order or keep the order imposed by FCFS. Let's analyze first the FCFS allocation as a (naive) competitive equilibrium without trading. A competitive equilibrium arises from each user maximizing his utility over the set of alternatives. In this case, the alternative choice set K for vehicle 1 are either decelerating ($k = 1$) or continuing at the same speed ($k = 0$):

$$\max_{k \in K} u_1(k, \theta_1) = \max_{k \in K} -\theta_1 d_1 k \tag{2.1}$$

The optimal allocation for agent 1 is $k = 0$ and the total social welfare is 0. However, the Pareto optimal allocation arises from the following program:

$$\max_{k \in K} u_1(k, \theta_1) + u_2(k, \theta_2) = \max_{k \in K} -\theta_1 d_1 k + \theta_2 d_2 (1 - k) \tag{2.2}$$

And such allocation is $k = 1$. However, agent 1 is worse off than before and would never accept this outcome without a compensation. An interesting way to address this issue is to provide vehicle 2 the right-of-way (the right to be externality-free). Vehicle 2 now proposes vehicle 1 a price p on the allocation $k = 1$. Vehicle 1 will accept in case $-\theta_1 d_1 - p \geq 0$, that is, if it is individually rational for vehicle 1. Several iterations of this take-it-or-leave-it bargaining procedure can be performed if necessary. Eventually, under complete information or non-strategic users, it can be proved that the system will converge towards the Pareto Optimal solution $(k^*, p^*) = (1, \theta_1 d_1)$ if agent 2 keeps making ascending offers to agent 1.

Alternatively, the right-of-way could be assigned to agent 1. Now, agent 1 could propose a quote p in exchange for slowing down. It is easy to check that the bargaining will converge again to a Pareto optimal solution, this time $(k^*, p^*) = (1, \theta_2 d_2)$. In any case, regardless of who has the right-of-way (property rights on it), the same degree of externality is selected. Figure 2.2 shows the bargaining trajectories, the green one corresponding to the first example and the red one corresponding to the second example. Other bargaining procedures would lead to the same allocation, although different prices may arise. This is related to the Coase theorem (Coase, 1960), a fundamental result in economics, which states that in the absence of transaction costs, efficiency is always guaranteed. Moreover, in case the income effects are small, which is certainly the case of quasilinear preferences, the outcome is independent of the assignment of property rights. This example is a clear case of P2P exchange of supply consumption: agents collaboratively consume supply by exchanging it on a decentralized manner.

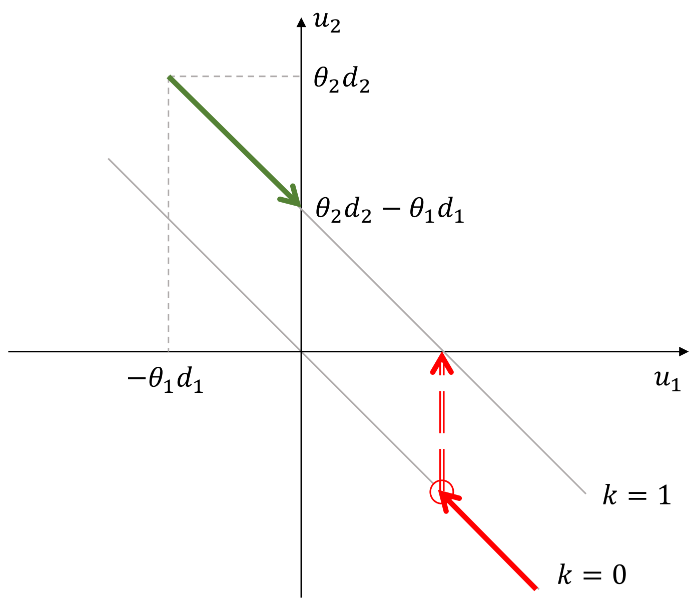


Figure 2.2: Pareto frontier ($k = 1$ line) with two different bargaining trajectories

These bargaining schemes have the advantage that their informational requirements are low. There is no need of a central authority to request or know users' preferences, although in fact, a third party or central authority is indirectly already present, actually replicated, in each of the vehicles' homologated communication and trading devices. Furthermore, these decentralized trading schemes suffer from a problem: if users are strategic about their private valuations, a Pareto-efficient allocation is not guaranteed (Myerson and Satterthwaite, 1983). However, by including a third party agent which is relied on by both agents, maximum expected efficiency can be guaranteed through a posted price mechanism, in which the third party proposes a guess price and parties have to agree. This idea is further developed in Chapter 6 of this dissertation.

Given the above results, it can be concluded that the actual assignment of property rights is not fundamental. In any case, more complex applications, e.g. traffic intersections, access queues or multimodal corridors, require infrastructure controllers. Infrastructure operators need to monitor traffic networks and transportation systems to evaluate their performance and to manage it directly in case there are special events or disruptions for construction and maintenance, as well as to ensure infrastructure resilience. In the rest of chapters, a centralized collaborative consumption approach will be used.

2.2 New Supply and Demand Framework

In this section, I provide a conceptual framework of transportation systems which considers collaborative consumption of supply. The structure of this framework is inspired by the work found in Florian and Gaudry (1980, 1983). Their objective in this well-accepted traditional framework was to describe the interaction between the different components of conventional transportation systems. Each of these components, labeled as "procedures", is a mapping which takes inputs and gives outputs as a result. This framework will serve as a starting point to embed one or many economic mechanisms which address the collaborative consumption of transportation supply. The work in this section is motivated by the need to view newer paradigms such as non-FCFS operations, that are possible now thanks to individual-level pricing, interactions and exchanges, examples of which follow in the next chapters.

Mechanisms are defined here in the most generic sense of the term: centralized control procedures which allocate transportation supply to a set of agents I and, at the same time, maximize one or many objectives from operators $j \in J$, which may be conflicting. In the case of multiple transport service operators, each operator $j \in J \backslash \{0\}$ will design its mechanism Z_j. I label infrastructure operator or regulator with $j = 0$. The mechanisms used in the applications found in this dissertation are direct revelation mechanisms which maximize economic efficiency while guaranteeing fairness and stability through envy-freeness. Nevertheless, it could also be the case of indirect revelation mechanisms, posted price mechanisms, markets, ascending auctions, etc. which could maximize revenue or any other objective.

In contrast to the traditional framework, our system considers users individually, instead of as an aggregate. With this in mind, we define the first procedure, demand D, as the cartesian product of individual actions D_i $\forall i \in I$. Each D_i takes as inputs each agent i's socioeconomic characteristics A_i, budget constraints B_i, user's history H_i, the current or last allocation X_i^* and price Π_i^*. Finally, we input mechanisms $\{Z_j^{(i)}\}$ to incorporate each user's knowledge of $\{Z_j\}$, which may not be complete. We use the accolades {} to denote a set. Each individual action procedure D_i outputs the "presence" of agent i and his message M_i, the latter applying to the case of participation of i in the collaborative system. This message can be understood as any kind of value information that the user communicates to the operator(s). Information not related to value, such as on vehicle speeds or positions, is already included in the agent's "presence" and not in the message M_i, even if they may be sent through the same communication protocol (e.g., DSRC in connected vehicle environments). Finally, each user's history H_i is the tuple of past allocations $\hat{X}_i$

and prices $\hat{\Pi}_i$ received, and messages $\hat{M}_i$ sent by user i:

$$[i, M_i] = D_i(A_i, B_i, X_i^*, \Pi_i^*, H_i, \{Z_j^{(i)}\}) \ \forall i \in I \tag{2.3}$$

$$H_i = (\hat{X}_i, \hat{\Pi}_i, \hat{M}_i) \ \forall i \in I \tag{2.4}$$

Since $D = \times_i D_i$, we can alternatively define $N = \times_i i$, $M = \times_i M_i$, $B = \times_i B_i$ $H = \times_i H_i$. The demand procedure then becomes:

$$[N, M] = D(A, B, X^*, \Pi^*, H, \{Z_j^{(I)}\}) \tag{2.5}$$

The second procedure is composed of each operator's supply actions S_j $\forall j \in J$. Each S_j outputs the mechanism $Z_j = (T_j, F_j)$ composed by two mappings T_j and F_j: the allocation and payment rules, respectively. Each operator is restricted to a regulation environment E, has an objective O_j (e.g., efficiency, revenue, ridership, accessibility); has resources R_j to manage the supply (e.g., personnel, fleet, funds transferred to agents) and observes or predicts the state of the system ST_j. j may also consider any knowledge of other operators' mechanisms $\{T_{-j}^{(j)}, F_{-j}^{(j)}\}$, where $-j \equiv J \backslash \{j\}$ and such knowledge may be incomplete.

$$[T_j, F_j] = S_j(O_j, E, R_j, ST_j, \{T_{-j}^{(j)}, F_{-j}^{(j)}\}) \ \forall j \in J \backslash \{0\} \tag{2.6}$$

$$ST_j = (\hat{X}_j, \hat{\Pi}_j, \hat{M}_j, \hat{O}_j) \ \forall j \in J \backslash \{0\} \tag{2.7}$$

Operator j's view of the state of the system ST_j is similar to users' histories. j knows the history of allocations $\hat{X}_j$, payments $\hat{\Pi}_j$, messages $\hat{M}_j$ of all users who messaged j and present agent set $\hat{N}$. In addition, j possess the history of objective evaluations $\hat{O}_j$. The infrastructure operator (or regulator) $j = 0$ produces regulation E, has resources R_0 (e.g., available funds for subsidy) and observes (perhaps partially) the system state ST_0, which does not include the messages.

$$E = S_0(O_0, \{ST_0\}) \tag{2.8}$$

$$ST_0 = (\{\hat{X}_j^{(0)}\}, \{\hat{\Pi}_j^{(0)}\}, \hat{O}_0, [\hat{N}]) \tag{2.9}$$

Between the demand and supply procedures, there is the exchange procedure G. This procedure is the actual collaborative consumption and exchange of supply within the transportation system. It can be understood as the exchange that is produced between the participating users and the operators. From a modeling perspective, it is the optimization problem that solves the mechanisms. G takes the set of agents present in the system N and the set of messages M sent to the operators, as demand-side input. It takes the set of mechanisms $\{T, F\}$ as supply-side input. It then outputs the resulting optimal allocations X^*, optimal payments Π^* and optimal objective values $O^* = \times_j O_j^*$:

$$[X^*, \Pi^*, O^*] = G([N, M], \{T_j, F_j\}) \tag{2.10}$$

This exchange procedure could be a collection of $|J|$ independent exchanges. An illustrative case is that of multiple Peer-to-Peer (P2P) ride-sharing companies which would include the ride exchange mechanism, as presented in Chapter 6 and in Masoud, Lloret-Batlle and Jayakrishnan (2017). Alternatively, the $|J|$ procedures could be linked to each other. That would be the case of multilateral exchanges between tour operators and users. Finally, it could of course be a case of a single exchange procedure if the supply side is controlled by a single infrastructure operator. This is the case of the queue-jumping policy and the exchange-based traffic control system PEXIC shown in the following sections. The framework above could eventually be extended to consider the activity location layer and cost/production layer of transport operators. To summarize:

Multiple operator case	Single (infrastructure) operator case
$[N,M] = D(A,B,X^*,\Pi^*,H,\{T_j^{(I)},F_j^{(I)}\})$	$[N,M] = D(A,B,X^*,\Pi^*,H,[T^{(I)},F^{(I)}])$
$[X^*,\Pi^*,O^*] = G([N,M],\{T_j,F_j\})$	$[X^*,\Pi^*,O^*] = G([N,M],[T,F])$
$[T_j,F_j] = S_j(O_j,E,R_j,ST_j,\{T_{-j}^{(j)},F_{-j}^{(j)}\})\ \forall j \in J\backslash\{0\}$	$[T,F] = S(O,E,R,ST)$
$E = S_0(O_0,\{ST_0\})$	$ST = (\hat{X},\hat{\Pi},\hat{O},[\hat{N},\hat{M}])$
$ST_j = (\hat{X}_j,\hat{\Pi}_j,\hat{M}_j,\hat{O}_j)\ \forall j \in J\backslash\{0\}$	$H = (\hat{X},\hat{\Pi},\hat{M})$
$ST_0 = (\{\hat{X}_j^{(0)}\},\{\hat{\Pi}_j^{(0)}\},\hat{O}_0,[\hat{N}])$	
$H = (\hat{X},\hat{\Pi},\hat{M})$	

Table 2.1: Procedures for transportation systems with collaborative consumption and exchange of supply

Here, $\{T_j^{(I)},F_j^{(I)}\}$ in the demand procedure indicates the set of individual knowledge items that every user $i \in I$ has, about mechanisms $\{T_j,T_j\}$.

Several types of equilibria can be defined in this framework. We could first focus on the partial demand-exchange equilibrium by fixing a mechanism and letting the demand levels and values equilibrate. Equivalently, a full demand-exchange-supply equilibrium could be defined. However, collaborative consumption of supply as an operational paradigm is a new idea in transportation systems, and thus there is a lack of behavioral data on the demand-side response to real-time exchange of supply. For this reason, this dissertation focus is in: first, presenting mechanisms as examples of supply functions; second, evaluating efficient and fair (envy-free, see 3.15) exchange procedures for fixed demand conditions with non-strategic agents; and third, doing a numerical evaluation as a sensitivity analysis of the policies' control parameters.

Last but not least, this new paradigm opens a new fourth type of problem: how to produce the regulatory environment E in the case of multiple operators. The regulator agency $j = 0$ needs to design and update the control environment E without fully observing the mechanisms i.e. not observing the prices or just observing some of the privacy-preserving allocations shared by the operators. The ramifications from the regulatory environment are myriad, and would directly make or break the kind of paradigms and operational schemes studied in the dissertation. As presented in the concluding thoughts of the introductory chapter, an elaborated study of this topic is beyond the scope of this dissertation.

Chapter 3

Methodology

This chapter presents the methodology used and developed in the dissertation. It includes a short introduction of mechanism design and social welfare concepts used and presents the new fundamental contributions in envy-freeness: **Dynamic Envy-freeness** and the **Constant Elasticity of Substitution Elasticity Intensity (CESEI) criterion**. The mechanism design introduction is based on (Mas-Colell et al., 1995), (Nisan and Ronen, 2001) and (Garg et al., 2008*a*,*b*). A general introduction to relevant social welfare concepts used, as well as the axiomatic framework for envy criteria are based on (Fleurbaey, 1994). The concept of envy is described in some detail in this chapter, albeit without making strong statements on the behavioral implications behind it. It is indeed a method to capture an individual's behavior in response to supply allocations to another individual or to collections of other individuals, for which the existing traveler behavioral models appear to be inadequate. A basic functional form for envy is considered for the purposes of this study, as a simple model is often the best for an initial study; however, elaborate functional forms with calibration schemes can be developed for modeling envy as a behavioral paradigm in the future. The chapter provides the economic principles and explains the usefulness of mechanism designs and envy-based models in the context of this dissertation.

3.1 Fundamental concepts

Definition 3.1. *Social choice function A social choice function is a function $f : \times_i \Theta_i \to X$, with which a social planner, mechanism designer or policy maker assigns a choice to each possible profile of agents' types $\theta = (\theta_1, \ldots, \theta_n) \in \times_i \Theta_i$.*

X is called the outcome set, and it is represented by a vector $(k, p_1, ..., p_n)$, where k is the resource allocation, i.e., objects, individual delays, and $p_1, \ldots, p_n$ are the prices charged to users. If p_i is positive, user i pays a price, if it is negative, he/she receives a payment.

The next step is to define an environment. An environment is a set of modeling assumptions that define a problem, an application or a set of them if the environment is general enough. From an environment, the designer selects the most appropriate social choice function, according to objectives and a particular implementation equilibrium.

Definition 3.2. *Environment*
An environment is the following set of assumptions:

- *A set I of n individuals who must make a choice from the outcome set X. The individuals will also be called agents or users indistinctively.*
- *Each agent i observes his parameter θ_i, known as type, from set Θ_i which determines his preferences.*
- *The vector of agents' types $\theta = (\theta_1, \ldots, \theta_n)$ is drawn from $\Theta = \times_i \Theta_i$ following a probability density function ϕ. This distribution does not need to be known by the designer.*
- *Each agent i attempts to maximize his utility function $u_i : X \times \Theta_i \Rightarrow \mathbb{R}$*

- *ϕ is assumed to be common knowledge. Also, given θ_i, the social planner and any agent can evaluate agent i's utility function.*

Given an environment, two questions arise:

1. Preference aggregation problem: given a type profile vector θ, which outcome $x \in X$ should be used?
2. Information Revelation Problem: assuming that the Preference Aggregation problem has been solved, how to elicit from the users their true types θ_i, which are private information?

To address these two problems, we define a mechanism:

Definition 3.3. *Mechanism*
A mechanism $M = ((S_i)_{i \in I}, g(\cdot))$ is a collection of action sets $(S_1, \ldots, S_n)$ and an outcome $g : S_1 \times \cdots \times S_n \to X$.

An option is to restrict the set of strategies to type revelation. Thus, $S_i = \Theta_i \forall i \in I$, and $g(\cdot) = f$. This produces a direct revelation mechanism:

Definition 3.4. *Direct revelation mechanism*
Given a social choice function $f : \Theta \to X$, a mechanism $D = ((\Theta_i)_{i \in I}, f(\cdot))$ is known as a direct revelation mechanism corresponding to $f(\cdot)$.

Mechanisms which are not direct revelation mechanisms are called indirect mechanisms. However, when we are concerned about implementing mechanisms with incentive compatibility, the revelation principle (Nisan et al., 2007) tells us that we can only focus on direct revelation mechanisms. This drastically diminishes the space of mechanisms to look at.

Alternatively, instead of asking users to reveal their types, the designer can just propose or *post* prices such that users accept or reject a particular outcome. This produces posted-price mechanisms. Posted-price mechanisms are useful since they always guarantee truthful elicitation of information. The problem these mechanisms face is: what is the price to set? The use of these mechanisms will be studied in Chapter 6.

We present next the environment that has been studied the most in mechanism design: the quasilinear environment. This environment is presented for introductory purposes, since the relevant theoretical results apply for this environment. A particular case of this environment, the linear environment, will be defined later. The utility function specification of the quasilinear environment is linear and separable on the price (also known as numeraire). From the part of the valuation $v_i(\cdot)$ it can take a general form. Agents can express their utility in terms of money, in such a way that there no welfare effects. This assumption holds valid in situations where the monetary transfers are relatively small with respect to the agents' incomes, such as in the applications explored in this dissertation.

Definition 3.5. *Quasi-linear environment*

$$X = \left\{ (k, p_1, \ldots, p_n) \mid k \in K, p_i \in \mathbb{R} \; \forall i = 1, \ldots, n \, , \sum_i p_i \geq 0 \right\} \tag{3.1}$$

A social choice function in this environment takes the form $f(\theta) = (k(\theta), p_1(\theta), \ldots, p_n(\theta))$. For a direct revelation mechanism $D = ((\Theta_i)_{i \in I}, f(\cdot))$. Every agent i takes of course a quasi-linear utility specification:

$$u_i(x, \theta_i) = u_i(k, p_1, \ldots, p_n, \theta_i) = v_i(k, \theta_i) - p_i + m_i \tag{3.2}$$

Where u_i is the utility function, v_i is the valuation function, that is, how much agent i values allocation k. p_i is the price charged to user i, which can be positive or negative and m_i is his or her initial endowment of money, which we will consider to be zero.

If $v_i(\cdot)$ can be defined as the product of the type θ_i, which expresses units of money per units of allocation times the "units of allocation" y_i, the environment is called **linear environment**. The types θ_i used in this dissertation will represent Value Of Time (monetary valuation of time spent in the transportation system), Value Of Delay Savings (monetary valuation of delay savings) or Value Of Distance (valuation of cost per distance suffered). The units of allocation will be either time or distance.

Definition 3.6. *(Allocative Efficiency, AE) A social choice function $f(\cdot) = (k(\cdot), p_1(\cdot), \ldots, p_n(\cdot))$ is allocative efficient if $\forall \theta \in \Theta$, $k(\theta)$ satisfies the following condition:*

$$k(\theta) \in \operatorname*{argmax}_{k \in K} \sum_{i=1}^{n} v_i(k, \theta_i) \tag{3.3}$$

Definition 3.7. *(Pareto Efficiency) A social choice function $f(\theta)$ is Pareto efficient (or also, Pareto optimal) if for every $f'(\theta) \neq f(\theta)$ and $\forall \theta = (\theta_1, ..., \theta_n)$*

$$u_i(f', \theta_i) > u_i(f, \theta_i) \Rightarrow \exists j \in I \mid u_j(f', \theta_j) < u_j(f, \theta_j) \tag{3.4}$$

Definition 3.8. *(Budget-Balancedness) A social choice function $f(\cdot) = (k(\cdot), p_1(\cdot), \ldots, p_n(\cdot))$ is (strictly) budget-balanced if $\forall \theta \in \Theta, p_1(\theta), ..., p_n(\theta)$ satisfies:*

$$\sum_{i \in I} p_i(\theta) = 0 \tag{3.5}$$

The social function f is said to be weakly budget-balanced when Definition 3.5 holds with $\geq$.

Theorem 3.1. *(AE + quasi-linear preferences + SBB = Pareto efficiency) A social choice function $f(\cdot) = (k(\cdot), p_1(\cdot), \ldots, p_n(\cdot))$ which is allocative-efficient, and strictly budget-balanced under quasi-linear preferences is Pareto efficient.*

Proof. (by contradiction): Suppose that the optimal allocation $(\boldsymbol{k}, \boldsymbol{p})$ is not Pareto optimal. Then all agents are better off under a new solution $(\boldsymbol{k}', \boldsymbol{p}')$. This new solution is also allocative-efficient and budget-balanced. Let $(w_i)_{i \in I} \in \mathbb{R}^{|I|}$ be the vector of individual valuation differences between both solutions such that $w_i = v_i(k(\boldsymbol{\theta}), \theta_i) - v_i(k'(\boldsymbol{\theta}), \theta_i)$. Let $(\rho_i)_{i \in I} \in \mathbb{R}^{|I|}$ be the vector of individual price differences between both solutions such that $\rho_i = p_i(\boldsymbol{\theta}) - p_i^{'}(\boldsymbol{\theta})$. Clearly, $\sum_i w_i = 0$ and $\sum_i \rho_i = 0$ since both solutions are budget balanced and allocative efficient. This means that at least one agent has to be worse off under $(\boldsymbol{k}^{'}, \boldsymbol{p}^{'})$, contradicting the fact that all $(\boldsymbol{k}, \boldsymbol{p})$ is not Pareto optimal. □

We view Pareto efficiency as the case when no user utility can be improved with no diminishing of anybody else's utility. Alternatively, we will also speak of a policy being Pareto Optimal with regard to a second policy when, regardless of, say, his/her vehicle's type v_i, an agent's utility is always larger under the first policy than under the second. In chapter 4, we will call a policy to be Pareto optimal with regard to another in an average sense. That is, if a particular type of agent has more utility on average, it is called Pareto optimal.

Besides truthfulness, a designer is interested in the users' willingness to participate in the mechanism, called individual rationality.

Definition 3.9. *(Ex-post Individual Rationality) A social choice function $f(\cdot) = (k(\cdot), p_1(\cdot), \ldots, p_n(\cdot))$ is Ex-Post Individual Rational (EPIR) if:*

$$v_i\Big(k(\theta_i, \theta_{-i}); \theta_i\Big) - p_i(\theta_i, \theta_{-i}) \geq \bar{u}_i(\theta_i) \forall i \in I, \forall \theta_i \in \Theta_i, \forall \theta_{-i} \in \Theta_{-i} \tag{3.6}$$

where $\bar{u}_i(.)$ is agent i's utility from not participating in the mechanism.

If EPIR is satisfied, an agent would be willing to truthfully participate in the mechanism rather than stay out.

3.2 Incentive Compatibility

We say that a mechanism M implements a social choice function $f(\cdot)$ when there is a pure strategy equilibrium $s^*(\cdot)$ of the Bayesian game induced by M such that $g(s^*(\theta)) = f(\theta), \forall\theta \in \Theta$.

The actual implementation will depend of the equilibrium flavor the designer wishes to induce on the agents. The two most common, and used in the literature, are dominant strategy implementation and Bayesian-Nash implementation.

Definition 3.10. *Weakly Dominant strategy equilibrium*
A pure strategy profile $s^d(\cdot)$ of the game induced by the mechanism M is said to be a weakly dominant strategy equilibrium if it satisfies the following condition.

$$\begin{aligned} &u_i(g(s_i^d(\theta_i), s_{-i}(\theta_{-i})), \theta_i) \geq u_i(g(s_i'(\theta_i), s_{-i}(\theta_{-i})), \theta_i) \\ &\forall i \in I, \forall \theta_i \in \Theta_i, \forall \theta_{-i} \in \Theta_{-i}, \forall s_i'(\cdot) \in S_i, \forall s_{-i}(\cdot) \in S_{-i} \end{aligned} \tag{3.7}$$

Definition 3.11. *Bayesian-Nash strategy equilibrium*
A pure strategy profile $s^b(\cdot)$ of the game induced by the mechanism M is said to be a Bayesian-Nash strategy equilibrium if it satisfies the following condition.

$$\begin{aligned} E_{\theta_{-i}}[u_i(g(s_i^b(\theta_i), s_{-i}^b(\theta_{-i})), \theta_i)|\theta_i] \geq E_{\theta_{-i}}[u_i(g(s_i'(\theta_i), s_{-i}^b(\theta_{-i})), \theta_i)|\theta_i] \\ \forall i \in I, \forall \theta_i \in \Theta_i, \forall s_i'(\cdot) \in S_i \end{aligned} \tag{3.8}$$

Ensuring incentive compatibility by eliciting users' parameters is often a difficult task. A common alternative is to design posted-price mechanisms. A posted-price mechanism is a mechanism where the designer fixes one or many prices which do not depend on users' parameters. Then, users observe those prices and decide to take a particular action. Posted-price mechanisms are also known as take-it-or-leave-it-price mechanisms.

How the designer decides which prices to set is generally derived from previous agents' behavior data. Incentive compatibility is trivially guaranteed since users are worse of from lying, which is equivalent to taking an action contrary to their best interests. However, there is an inherent loss in efficiency. In the context of bilateral trading, a price set too high could prevent a possible trade from occurring, leaving the item to the agent who values it the least. Posted-price mechanisms are used in Chapter 6 in the context of dynamic ridesharing.

3.2.1 Dominant Strategy Implementation

A mechanism M implements the social choice function $f(\cdot)$ in dominant strategy equilibrium if there is a weakly dominant strategy equilibrium s^d of the Bayesian game induced by M such that $g(s^d(\theta)) = f(\theta)\ \forall \theta \in \Theta$.

Definition 3.12. *Dominant Strategy Incentive Compatibility (DSIC)*
The SCF $f(\cdot)$ is dominant-strategy incentive-compatible or truthfully-implementable in dominant strategies whenever:

$$u_i(f(\theta_i,\theta_{-i}),\theta_i) \geq u_i(f(\theta_i',\theta_{-i}),\theta_i), \forall i \in I, \forall \theta_i \in \Theta_i, \forall \theta_{-i} \in \Theta_{-i}, \forall \theta_i' \in \Theta_i \tag{3.9}$$

that is, eliciting the type truthfully is a dominant strategy equilibrium of the game induced by the direct mechanism D

In a quasi-linear environment, Groves (1973), showed that there exists a family of social choice functions which are both allocatively efficient and Dominant Strategy Incentive Compatible (DSIC)

Theorem 3.2. *Groves theorem: Let $f(\cdot) = (k^*(\cdot), p_1(\cdot), \ldots, p_n(\cdot))$ be a social function which is allocative efficient. f can be dominant-strategy incentive compatible its payment function follows* *this scheme:*

$$p_i(\theta) = h_i(\theta_{-i}) - \left[\sum_{j \neq i} v_j(k^*(\theta), \theta_j)\right] \forall i \in I \tag{3.10}$$

Where $h_i(\cdot)$ is an arbitrary function of θ_{-i} up to satisfying the feasibility condition $\sum_i t_i(\theta) \geq 0 \; \forall \theta \in \Theta$

This theorem leads to the following family of direct revelation mechanisms:

Definition 3.13. *(Groves mechanism) A direct revelation mechanism $D = ((\Theta_i)_{i \in I}, f(\cdot))$ which satisfies 3.3 and 3.10 is known as Groves mechanism.*

There is a special case of Groves mechanism, discovered by Clarke in Clarke (1971), known as the Clarke or pivotal mechanism.

Definition 3.14. *(Vickrey Clarke Groves mechanism)*

$$k^* = \underset{k \in K}{\operatorname{argmax}} \sum_{i \in I} v_i(k(\theta)) \tag{3.11}$$

$$p_i(\theta) = \sum_{j \neq i} v_j(k^*_{-i}(\theta_{-i})) - \sum_{j \neq i} v_j(k^*(\theta), \theta_j); \forall i \in I \tag{3.12}$$

Due to Groves theorem, we know that the VCG mechanism is allocative-efficient and DSIC. What about Individual Rationality (IR), or Budget-Balancedness (BB)? There are positive results for both constraint sets, as long as the environment satisfies additional properties.

Definition 3.15. *Choice set monotonicity*

A mechanism M is choice set monotone if X weakly increases as additional agents are introduced into the system: $K_{-i} \subset K \; \forall i \in I$.

Definition 3.16. *No negative externality*

Let M be a choice set monotone mechanism. M has no negative externality if $\forall i \in I, \theta \in \Theta, k^*_{-i}(\theta_{-i}) \in B^*(\theta_{-i}) : v_i(k^*_{-i}(\theta_{-i}), \theta_i) \geq 0$.

Proposition 3.1. *Individual Rationality of VCG mechanisms*

Let M be a Clarke mechanism such that:

1. $\bar{u}_i(\theta_i) = 0 \ \forall \theta_i \in Theta_i; \forall i \in I$
2. *is choice set monotone*
3. *satisfies no-negative externality property*

Then, M is Individually Rational.

Proof. See Mas-Colell et al. (1995) □

Concerning budget balancing, there is another possibility result which involves an additional property:

Definition 3.17. *No single agent effect*

Mechanism M has no single agent effect if $\forall i \in I, \theta \in \Theta, k^*(\theta) \in B^*(\theta)$, there is a $k \in K_{-i}$ such that:

$$\sum_{j \neq i} v_j(k, \theta_j) \geq \sum_{j \neq i} v_j(k^*(\theta), \theta_j) \tag{3.13}$$

Proposition 3.2. *If a VCG mechanism satisfies the no single agent effect property, then each transfer should be non-positive. Thus, weakly budget balance is satisfied.*

Proof. See Mas-Colell et al. (1995) □

These results will prove useful later in Chapter 7.

3.2.2 Bayesian-Nash implementation

This is a weaker concept than a Dominant Strategy IC (DSIC), and is used in Chapter 4

Definition 3.18. *Linear environment:*

1. *Each type* θ_i *lies in* $\Theta_i = \left[\underline{\theta}_i, \overline{\theta}_i\right]$ $\forall i \in \{1,\ldots,n\}$, *where* $\underline{\theta}_i < \overline{\theta}_i$.
2. *Agents' types are statistically independent*$\phi(\cdot) = \times_i \phi_i(\cdot)$.
3. $\phi_i(\theta_i) > 0 \quad \forall \theta_i \in \left[\underline{\theta}_i, \overline{\theta}_i\right]$ $\forall i \in \{1,\ldots,n\}$
4. *Each agent i's utility function has the form:* $u_i(x,\theta_i) = \theta_i v_i(k) + t_i$

Theorem 3.3. *Myerson's Characterization Theorem) In a linear environment, a social choice function* $f() = (k(), p_1(), \ldots, p_n())$ *is BIC if and only if,* $\forall i \in \{1,\ldots,n\}$:

1. $\overline{v}_i(\cdot)$ *is non-decreasing.*
2. $U_i(\theta_i) = U_i(\underline{\theta}_i) + \int_{\underline{\theta}_i}^{\theta_i} \overline{v}_i(s)ds$ $\forall \theta_i \in \Theta_i$

Where $\overline{v}_i(s)$ *is the average valuation function for agent i when he/she reveals type s.* $U_i(\theta_i)$ *is the expected utility of user i when he/she has type* θ_i *and all agents reveal their types truthfully, including i.*

By replacing the utility function specification into the second condition keeping the expectation, we can isolate the average transfer $\bar{p}_i(\theta_i)$.

$$\overline{p}_i(\theta_i) = \overline{p}_i(\underline{\theta_i}) - \underline{\theta_i}\overline{v_i}(\underline{\theta_i}) + \theta_i \overline{v_i}(\theta_i) - \int_{\underline{\theta_i}}^{\theta_i} \overline{v_i}(\theta_i)ds \tag{3.14}$$

$\overline{p}_i(\theta_i)$ describes a family of price functions, with a free constant $\overline{p}_i(\underline{\theta_i})$.

3.3 Fairness under envy-based user behavior

Mechanism design can also focus on implementing equilibria with goals other than incentive compatibility. For instance, assuming that agents elicit their private information truthfully, a designer might be concerned whether users are satisfied with their current allocations and prices, or whether some of them would be better off by exchanging allocations. In other words, the designer needs a model for the users response to the allocations to them and others. The natural tendency in a transportation researcher would be to rely on the various behavioral mechanisms studied in transportation demand modeling in the past.

These models were developed under the assumption that the users were aware only of the supply environment, or the supply-performance environment resulting from the collective behavior of others, not of data at individual level from other users. For the purposes of this dissertation, however, what is important is any user's response to the supply allocated to other individual users, or to coalitions formed among other individuals. The existing models would appear largely inadequate for this purpose.

Since data to calibrate models of such response will become available in the future, we can indeed contemplate different kinds of behavioral models for individuals' response to other individuals (one-to-one or one-to-a-known-collection). It is in this context that envy models are proposed here as a suitable alternative. Simply put, ”envy” refers to how each user evaluates, with his on her own yardsticks, what another user is allocated, in comparison with his or her own allocation. These concepts have existed in economic welfare theory but have not been viewed as a plausible behavioral mechanism, as done in this dissertation.

Under the assumption that users are utility maximizers, for a given price or cost vector, they would choose an allocation that maximizes their utilities. Envy-freeness actually captures such situations as well, since it ensures that no user is willing to swap their allocations at their current prices. The key aspect to note, is that the envy that any agent feels is a function of his or her own valuation of others' allocations, which would normally be different than the valuation done by the other agents for themselves. This could be under the presence of a central agent whose goal is to maximize a particular objective such as social welfare or revenue. Alternatively, one could allow users to trade with each other the supply of transportation that they each possess, in a decentralized fashion. These trades would be motivated by envy between agents as well and could eventually increase global social welfare. Nevertheless, the spatio-temporal nature of transportation systems makes distributed approaches difficult to implement in practice than a more centralized allocative scheme. Thus, this dissertation will focus on a centralized approach.

It is important to note that I do not consider "envy" to be the same or similar as "jealousy". To clarify, I do not refer to "envy" as a negative externality experienced by an agent due to someone else's allocation, but as the desirability of that allocation with regard to the current one. This is why an agent here is said to be"willing" to trade away the current allocation, as opposed to "wanting" to change it. Kolm (1995) presents a discussion on this finer point.

Envy-freeness started as a problem of fair division, with cake-cutting (Brams and Taylor, 1996) as a classic example, in which divisible goods have to be split fairly among agents; these are the cases of land partitioning or divorce settlements. Economists showed that in the presence of a divisible good such as money, an envy-free Pareto efficient allocation always exists Alkan et al. (1991). On the other hand, computer scientists explored the idea of profit maximizing envy-free pricing. The idea behind this is that, under envy-free pricing, both buyers and sellers maximize their utilities, and therefore do not have incentives to switch allocations. This pricing scheme also has proven useful as a benchmark for auctions (Hartline and Yan, 2011; Guruswami et al., 2005).

3.3.1 Static envy-freeness

To describe envy-freeness, an environment needs to be described first. Let I be the set of users and X the space of feasible allocations. Each user has a linear utility function: $u_i = v_i x_i - p_i$ where v_i is i's valuation or type, that is, how much user i values one unit of allocation x_i, and p_i is the price paid (or received) for x_i. Let $\mathbf{v}$ be the vector of all user valuations. $\mathbf{x}$ and $\mathbf{p}$ are the allocation and price vectors that includes all agents. In transportation problems, we would normally consider the allocation to be delays, travel time savings, orderings in a queue, etc. Note that the agents may indeed prefer less allocations, such as for instance in the case of delay. Valuations are the monetization of these allocations based on the willingness to pay for them. We say that a resource allocation $A = (\mathbf{x}, \mathbf{p})$ is envy-free when no user would be better off from exchanging allocations with another one at the current prices. Based on the valuations each agent has, his/her valuation of any other agent does not induce envy, in other words. This is equivalent to saying that A maximizes utility for each user. User optimality and allocation stability gives this concept a fair and equitable interpretation. This is represented mathematically by the following condition:

$$v_i x_i - p_i \geq v_i x_j - p_j \quad \forall i, j \in I, i \neq j \tag{3.15}$$

It is easy to prove that the well-known Wardrop principle for user equilibrium in the traditional formulation of the static network traffic assignment problem results in an envy-free solution. This is because the complementary slackness and first order conditions are in fact in the form of (3.15). Envy-freeness, however, applies to individuals, and not at the route flow level as in the traditional treatment of the traffic assignment problem. In general, frameworks that ensure individual level comparisons can be expected to have higher information requirements, but that is not the case when it is based on envy-freeness. In fact, one of the advantages of testing for envy-freeness is that it does not involve interpersonal utility comparisons. No agent knows the valuation of another agent, and each agent only uses his/her own valuation of the other agent in any comparison. However, envy-freeness is a very strong concept and in the presence of certain constraints it may not even exist. In these cases, envy-freeness can be relaxed in several ways. From a behavioral sense, "relaxing envy-freeness" actually implies an assumption that users may agree to not being perfectly envy-free, which in turn is akin to assuming a boundedly-rational behavior in them.

Generally, in transportation problems, users have unit demands. That is, they desire just one unit of resource, since each user can only be on one trip at a time. This unit is normally an assigned path, a delay assignment or a position in a queue. In situations where there is no possibility of price discrimination, or when a set of users receive the same allocation, envy-free pricing assigns the same price to all the users in such a group. I proceed to characterize envy-free allocation and prices in the defined environment:

Theorem 3.4. *Suppose* $\boldsymbol{v}$ *is ordered decreasingly. An outcome* $(\mathbf{x}, \mathbf{p})$ *is envy-free iff:*
a) The allocation is monotonous decreasing:

$$x_1 \geq \ldots \geq x_n \tag{3.16}$$

b) The prices follow from this expression

$$\forall i \in I,\ p_i = \sum_{k \leq i} \alpha_k,\ \alpha_1 \in \Re,\ \alpha_k \in [v_{k-1}(x_k - x_{k-1}), v_k(x_k - x_{k-1})]\ \forall k > 1 \tag{3.17}$$

Condition b) starts with the first agent WLOG.

Proof. *EF* $\Rightarrow$ *monotonicity*: $v_i x_i - p_i \geq v_i x_j - p_j$ and $v_j x_j - p_j \geq v_j x_i - p_i$. By reordering these two envy relationships and adding them we get the monotonicity condition $(x_i - x_j)(v_i - v_j) \geq 0$.

$EF \Rightarrow$(3.17): Set $p_1 = \alpha_1$. By the EF conditions we have $v_2(x_2 - x_1) \geq p_2 - \alpha_1 \geq v_1(x_2 - x_1)$. By adding $\alpha_1$to both sides of the inequality we get $v_2(x_2 - x_1) + \alpha_1 \geq p_2 \geq v_1(x_2 - x_1) + \alpha_1$. Subsequently, for $i < j$: $v_j(x_j - x_i) + \sum_{k \leq i} \alpha_k \geq p_j \geq v_i(x_j - x_i) + \sum_{k \leq i} \alpha_k$.

(3.16) and (3.17)$\Rightarrow$ EF: Suppose $i < j$. $v_i x_i - p_i = v_i x_j + v_i(x_i - x_j) - \sum_{k \leq i} \alpha_k = v_i x_j + \sum_{k=1}^{j-1} v_i(x_k - x_{k+1}) - \sum_{k \leq i} \alpha_k \geq v_i x_j + \sum_{i < k \leq j} v_{k+1}(x_{k+1} - x_k) - \sum_{k \leq i} \alpha_k \geq v_i x_j - \sum_{k \leq i} \alpha_k - \sum_{i < k \leq j} \alpha_k = v_i x_j - p_j$. The same reasoning can be used analogously for $j < i$. □

The term α_1 is found by imposing an additional condition. For instance, we would be interested in imposing budget balance, that is, $\sum_i p_i = 0$. Given a monotonically decreasing allocation $\mathbf{x}$, we find all the α_i $\forall i > 1$ using (3.17). Then, we find α_1 by imposing the budget balanced condition: $\alpha_1 = \frac{1}{n} \sum_{i>1}(n - i + 1)\alpha_i$ which allows us to build the envy-free price vector $\mathbf{p}$. Alternatively, when agents' allocations are always non-negative, we could impose individual rationality for one of the agents, i.e $\alpha_1 \in [0, v_1 x_1]$. In this particular case, selecting one of the interval extremes for each agent would give us revenue minimizing or revenue maximizing prices, respectively.

Finally, envy relations between agents can be represented by envy graphs, in which edges represent the presence or intensity of envy from one agent to another. This is useful to establish restrictions on envy relations, since strict envy-freeness is a very strong condition that generally does not hold in many situations. For instance, it could be of interest to restrict envy on the basis of the arrival order of agents or actual connectivity between agents: if agents are rather disconnected (which could mean large spatio-temporal distances between them in a transportation context), it does not make sense to model envy relations between them. This idea has been explored in Endriss et al. (2006); Chevaleyre et al. (2007) for distributed trades in exchange economies. While such distributed exchanges can be considered in our problems as well, given the spatio-temporal complexity of transportation systems and the complementarity in preferences of transportation users, we believe that centralized approaches are more operationally efficient in terms of communication complexity for the kind of problems we consider.

3.3.2 Dynamic envy-freeness

Surprisingly, envy-freeness for dynamic situations has never been explored in economic literature. The time dimension being fundamental in transportation applications, we propose the concept of dynamic envy-freeness. Informally, it could be defined as a succession of states in which no agent i envies any other agent j at any time instant t, given the current allocation x_i^t, current price p_i^t and accumulated prices until that time instant, π_i^{t-1}:

Definition 3.19. *Dynamic envy-freeness:*

$$v_i x_i^t - p_i^t - \pi_i^{t-1} \geq v_i x_j^t - p_j^t - \pi_j^{t-1} \ \forall i, j \in I, i \neq j, \forall t \in T \tag{3.18}$$

In contrast to standard resource allocation, where indivisible items are transferable at any time instant, in transportation problems the allocations are experienced by each agent and therefore non-transferable, such as delays. x_i^t are thus the predicted accumulated delays at time t: $x_i^t = t_i^{dep} - a_i$, where a_i are the arrival times into the system and t_i^{dep} are the departure times from the system. This new concept is developed further in chapter 4.

The real-time dimension of traffic signal control requires the theory of envy-freeness to be extended to the time dimension. The question arise on: (1) how the envy is instantaneously felt among agents, (2) to what degree of envy an agent experiences at a particular moment in time, and (3) how it should be measured. We proceed to address these concerns in defining the environment for dynamic envy freeness.

We first define states $s_t \in S_t, \forall t \in \Re^+$. Each state $s_t = (\boldsymbol{i}_t, \boldsymbol{d}_t, \boldsymbol{p}_t, \boldsymbol{\pi}_t) \in I_t \times \Re^+ \times \Re \times \Re$ describes the set of possible agents $\boldsymbol{i}_t$, predicted accumulated delays $\boldsymbol{d}_t$, prices $\boldsymbol{p}_t$ and accumulated prices $\boldsymbol{\pi}_t$ during instant t. Suppose that at instant t, each agent $i \in \boldsymbol{i}_t$ experiences an excess envy ϵ_{it}. These envy values only change once the state changes. Since it is assumed that a predicted vehicle arrival time at the intersection does not change after the vehicle is detected, the only two elements that can alter envy levels are changes in prices and departure times. This only happens once new vehicles enter the system or some present vehicles leave it. In conclusion, dynamic envy-freeness must satisfy the condition that if a state stays invariant over a period of time, the individual excess envies should remain constant.

Definition 3.20. *Constant envy to state invariance;*$\forall t \in \Re^+$, *if* $s_t = s_{t+1}$, *then* $\epsilon_{it} = \epsilon_{i,t+1} \forall i \in I_t \cap I_{t+1}$.

This postulate has two desirable consequences for rolling-horizon dynamic optimization. Suppose that t refers to the time points at which the system is updated or optimizations are executed. Specifying the envy-free conditions as a function of accumulated predicted delays and accumulated prices, satisfies the property above, and makes the satisfaction of envy-free conditions independent from the frequency of optimization or time granularity. In particular, in the application of traffic signal control, if we wished to optimize the phasing more frequently and states remained invariant, the envy levels would remain constant. Secondly, it also prevents double counting in prices for agents who are present in multiple periods, despite the rolling horizons being overlapped. Thus it avoids the nonsensical case of the envy calculations being dependent on the modeling parameters selected by the analyst. For this problem the (relaxed) dynamic envy-free conditions for every rolling horizon optimization h are:

$$\epsilon_{ij} - \theta_i d_i - p_{i,h} - \pi_{i,h-1} \geq -\theta_i d_j - p_{j,h} - \pi_{j,h-1} \; \forall i,j \in I_h, i \neq j \tag{3.19}$$

3.3.3 Constant Elasticity of Substitution Envy Intensity (CESEI) criterion

Envy-freeness relaxation requires an objective function criterion to rank the slack envy terms ϵ_{ij}. Furthermore, the problem is now dynamic, and thus the future envy conditions at a given time depend on previous periods' allocations and prices. Existing criterion, such as minimax envy intensity Diamantaras and Thomson (1990), envious and envied intensity Fleurbaey (2008) are based on the minimax operator. The use of minimax is founded on the understanding of the envy terms ϵ_{ij} as a quantity of resource which has to be transferred or subtracted to agent i to not to make him feel envious of agent j. Therefore, for every agent i the only ϵ_{ij} term that matters is the largest one. However, this approach has a drawback: it overlooks all the envy relations that are not binding.

We provide next three reasons in which considering more envy terms is more beneficial. First, there are applications in which values and delays are not correlated, such as traffic signal control (This lack of correlation is due to the arrival order being independent of the individuals' type.) Secondly, agents are present during multiple periods and carry accumulated payments. Thirdly, value of time functions and in particular the value of delay savings function used in chapter 4 are highly skewed: there are many low valuation agents that need to be compensated by very few high valuation vehicles. Thus, new criteria are needed which account for more envy relations and compensates more uniformly across the whole valuation range. We define next a family of criteria that addresses these desired characteristics.

Definition 3.21. *Constant elasticity of substitution envy intensity criterion (CESEI)*

$\forall e = (\boldsymbol{v}, \boldsymbol{d}) \in E, \forall \mathbf{p} \in P, \beta_j \geq 0\ \forall j \in I_h, \rho \leq 1, \rho \neq 0, \gamma_i \geq 0 \mathbf{p} \in S_{CESEI}(e) \Leftrightarrow \forall \mathbf{p}' \in P,$:

$$\sum_{i \in I_h} \gamma_i \left(\sum_{j \neq i} \left(\beta_j \epsilon_{ij}(\mathbf{p}) \right)^{\rho} \right)^{1/\rho} \leq \sum_{i \in I_h} \gamma_i \left(\sum_{j \neq i} \left(\beta_j \epsilon_{ij}(\mathbf{p}') \right)^{\rho} \right)^{1/\rho}, \epsilon_{ij} \geq 0\ \forall i, j \in I_h \tag{3.20}$$

Here, S_{CESEI} is the solution set that satisfies CESEI criterion. The excess envy terms in the CESEI criterion acquire a different interpretation. In the existing criteria, the minimax operator selects the excess envy such that an agent does not feel envious of any other agent. Since all the envy relations are considered now, envy intensities acquire a different meaning: agents are envious of every other agent j, at some degree $1/\beta_j$. In line with the CES expression for production and utility functions found in microeconomics, CESEI criterion expression achieves especial forms found in the literature. In particular, when $\rho = \infty$ CESEI criterion resembles the Envious Intensity criterion found in Fleurbaey (2008). When $\rho = 1$ the criteria is identical to the Weighted Additive Envy Intensity criterion used in chapter 4. When $\rho = 1$, it can be argued that there is perfect substitution between the envy terms. One must be careful though, that we are considering $\rho >= 1$, while in consumer theory, elasticity of substitution is defined for $\rho <= 1$. Finally, for $1 < \rho < +\infty$, the following alternative CES expression is easier to compute, especially when $\rho = 2$, where CESEI optimization problem can be formulated as a quadratic program:

$$\mathbf{p} \in S_{CESEI}(e) \Leftrightarrow \forall \mathbf{p}' \in P,$$

$$\sum_{i \in I_h} \gamma_i \sum_{j \neq i} \left(\frac{1}{\rho} \right) \left(\beta_j \epsilon_{ij}(\mathbf{p}) \right)^{\rho} \leq \sum_{i \in I_h} \gamma_i \sum_{j \neq i} \left(\frac{1}{\rho} \right) \left(\beta_j \epsilon_{ij}(\mathbf{p}') \right)^{\rho}, \epsilon_{ij} \geq 0\ \forall i, j \in I_h \tag{3.21}$$

Minimizing the above expression is equivalent to minimizing (3.20), since the transformation from (3.21) to (3.20), shortly (3.20) = $(\rho * (3.21))^{(1/\rho)}$ is monotonic. We next show a few properties of the new CESEI criteria family.

Proposition 3.3. *The CESEI envy intensity criterion selects the set of envy free allocations whenever it is non empty.*

Proof. $\forall \mathbf{p} \in \mathbf{P}, \forall i \in I_h, \sum_{j \neq i} \left(\beta_j \epsilon_{ij}\right)^{\rho} > 0 \Leftrightarrow i$ envies some agent j, since $\epsilon_{ij} \geq 0 \; \forall i, j \in I_h$. Therefore, $\sum_{i \in I_h} \gamma_i \left(\sum_{j \neq i} \left(\beta_j \epsilon_{ij}(\mathbf{p})\right)^{\rho}\right)^{1/\rho} \Leftrightarrow \mathbf{p}$ is envy-free. □

This result is useful to prove that this new criterion satisfies two (relabeled here) axioms from Fleurbaey and Maniquet (2011): Equal payment for uniform allocation (EPUA) and Equal Utility for Uniform Type (EUUT).

Proposition 3.4. *The CESEI criterion satisfies* $\forall e \in E, \forall p \in P, \forall i, j \in I \mid d_i = d_j : \; p_i = p_j$ *(Equal Payment for Uniform Allocation, EPUA)*

Proof. When $d_i = d_j \; \forall i, j \in I$, the only prices possible are $\mathbf{p} = \mathbf{0}$, which is the envy-free allocation. By proposition 3.3, such envy-free allocation is selected, which completes the proof. □

Proposition 3.5. *The CESEI criterion satisfies* $\forall e \in E, \forall p \in P,$ *if* $\forall i, j \in I, \theta_i = \theta_j : \; \forall i, j \in I, u(p_i, d_i, \theta_i) = u(p_j, d_j, \theta_j)$ *(Equal Utility for Uniform Type, EUUT)*

Proof. Analogously, the only allocation that equates utilities when two individuals' types are the same is the envy-free allocation. By Proposition 3.3, such envy-free allocation is selected. □

The axioms and properties defined above are satisfied when we consider $\mathbf{p}^t + \boldsymbol{\pi}^{\mathbf{t}-1}$ as the prices to optimize.

3.4 Cooperative Game Theory: Coalitional Stability

This section provides a very short introduction to the main cooperative game theory concepts necessary to understand this dissertation. The reader is encouraged to access the following textbook (Peleg and Sudhölter, 2007) for a full, comprehensive introduction on cooperative game theory. The references cited in this document will clarify the reader on the advanced, less known concepts not covered in standard textbooks.

3.4.1 Characteristic Function Games (CFG)

A mechanism designer may also be interested in a particular outcome being stable from a coalitional point of view. Consider a group of agents N. Agents $i \in N$ can form coalitions and decide over an action, outcome to eventually split some profits or utility x_i. This magnitude is also called imputation. The value generated by coalition $S \subseteq N$ is $v(S)$. The function $v : 2^N \to \mathbf{R}$ is called the characteristic function. The pair $< N, v >$ defines a cooperative game in characteristic function form. Analogously, cooperative games can be expressed in terms of a cost function. This defines cost sharing games $< N, c >$: players no longer share profits but costs.

The designer is interested in knowing if all the agents could agree on a particular agent or outcome, in the sense that there is no subgroup of agents $S \subset N$ who is better off by breaking the the global agreement and going on their own. If such coalition S exists, they are called blocking coalitions. The coalition with maximum agreement (the whole set of players N) is called grand coalition. The pair $< N, v >$ is the cooperative game. It is easy to see that the definitions above define constraints and a feasible set. From these constraints emanates the fundamental concept of cooperative game theory: the core.

The core is the set of feasible payoffs (imputations) satisfying three properties: budget balance (also called efficiency), coalitional stability, which boils down to individual rationality when the coalitions are singletons:

Definition 3.22. *The core of a characteristic function game* $< N, v >$ *is:*

$$Core(N,v) := \{\boldsymbol{x} \in \Re^{+} \mid \sum_{i \in N} x_i = v(N),\ \sum_{i \in S} x_i \geq v(S)\ \forall S \subseteq N\} \tag{3.22}$$

As per usual in concepts based on feasibility, the question that now arises is about the non-emptiness of such program. The Shapley-Bondareva theorem addresses this question. Some definitions need to be given first:

Definition 3.23. *A map* $\lambda :\ 2^N \backslash \emptyset \rightarrow \Re^{+}$*is balanced iff* $\sum_{C \subseteq N} \lambda(C)\chi_C = \chi_N$.

Definition 3.24. *A game is balanced iff for every balanced map*λ*:*
$\sum_{C \subseteq N, C \neq \emptyset} \lambda(C) v(C) \leq v(N)$.

Theorem 3.5. *(Bondareva-Shapley)* $Core(N,v) \neq \emptyset \iff < N, v >$ is balanced.

Proof. :

$$(LP) \quad \begin{cases} \min\ x(N) \\ s.t.\ x(C) \geq v(C)\ \forall C \subseteq N, S \neq \emptyset \end{cases}$$

$$(DLP) \quad \begin{cases} \max\ \sum_{C \subseteq N} y_C v(C) \\ s.t. \begin{cases} \sum_{C \subseteq N} y_C \chi_C = \chi_N \\ y_C \geq 0\ C \subseteq N, C \neq \emptyset \end{cases} \end{cases}$$

From the complementary slackness conditions, $v(N) \geq \sum_{C \subseteq N} y_C v(C)\ \forall (y_C)_{C \subseteq N}$ is feasible.

□

Cooperative games $< N, v >$ can be classified in different categories in accordance with the properties of their characteristic function v.

Definition 3.25. *A cooperative game* $< N, v >$ *is* ***convex*** *iff:*

$$\forall i \in N,\ \forall S \subseteq T \subseteq N \backslash \{i\}, \quad v(T \cup \{i\}) - v(T) \geq v(S \cup \{i\}) - v(S) \tag{3.23}$$

If the game is a cost game $< N, c >$ then:

$$\forall i \in N,\ \forall S, T \subset N \backslash \{i\} \mid S \subset T\ c(T \cup \{i\}) - c(T) \leq c(S \cup \{i\}) - c(S) \tag{3.24}$$

Convexity is a sufficient condition for balancedness:

Theorem 3.6. $< N, v >$ *is convex* $\Rightarrow Core(N, v) \neq \emptyset$.

A superset of convex games is superadditive games.

Definition 3.26. *A cooperative game* $< N, v >$ *is superadditive if:* $\forall S, T \in N \mid S \cap T = \emptyset,\ v(S) + v(T) \leq v(S \cup T)$.

However, it is important to note that superadditivity does not guarantee balancedness.

Concepts based on feasibility regions have two inconvenient issues: they can be non-empty or they can contain an infinite number of solution points. Shapley (1953) defines a value which satisfies desirable axioms, that is unique and always exist. Unfortunately, it does not guarantee core-stability.

Definition 3.27. *Marginal contribution of agent* i *for coalition* $C \subseteq N \backslash \{i\}$ *is* $mc_i(C) = v(C \cup \{i\}) - v(C)$.

Let σ be a permutation on N, which indicates the order of joining a coalition. The marginal contribution arising from such a permutation is: $mc(\sigma) \in \Re^n$ and agent i obtains $mc_i(\{\sigma(j) \mid j < i\})$.

Definition 3.28. *The Shapley value of a cooperative game $< N, v >$ is:*

$$\phi(v) = \frac{\sum_{\sigma \in \Pi(N)} mc(\sigma)}{n!} \tag{3.25}$$

This value is in fact the average marginal contribution over all the permutations.

The following expression for the Shapley value is more efficient,

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!} (v(S \cup \{i\}) - v(S)) \ \forall i \in N \tag{3.26}$$

Properties:

- Efficiency: the total gain is distributed $\sum_{i \in N} \phi_i(v) = v(N)$
- Symmetry: agents i, j are symmetric if $\forall S \subset N \mid i, j \notin S,\ v(S \cup i) = v(S \cup j)$. If agents i, j are symmetric $\Rightarrow \phi_i(v) = \phi_j(v)$
- Dummy player: an agent is null if $v(S \cup i) = v(S)\ \forall i \in N$
- Additivity: If two coalition games $< N, v >$ and $< N, w >$ are combined, $\phi_i(v + w) = \phi_i(v) + \phi_i(w)$

The Shapley value may not be in the core, even if the core is non-empty. However,

Theorem 3.7. *The Shapley value of a superadditive game is individually rational*

3.4.2 Partition Function Games (PFG)

The above section showed the main results for cooperative games in characteristic function form. However, these games are not expressive enough to incorporate coalitional externalities. Literature has defined a new class of games, from which CFG are a particular case, which can express coalitional externalities. There is again a set of players N and now a valuation function $v : 2^N \times \mathcal{P} \to \mathbf{R}$. This defines a game $< N, v >$ expressed in Partition Function Form. Every coalition S will have different payoffs based on the complementary coalitions $T \in \mathcal{P} - S$. This can capture coalitional externalities, that is, changes in $v(C, \cdot)$ due to other coalitions $S, T \in \mathcal{P} - C$ merging or splitting.

Coalitional externalities can be either positive or negative:

Definition 3.29. *Positive externalities:*

$$\begin{gathered}\forall C, S, T \mid C \cap S \cap T = \emptyset \text{ and } \forall \rho \in \mathcal{P}(N - (S \cup T \cup C)) : \\ v(C; \{S \cup T, C\} \cup \rho) > v(C; \{S, T, C\} \cup \rho)\end{gathered} \tag{3.27}$$

Definition 3.30. *Negative externalities:*

$$\begin{gathered}\forall C, S, T \mid C \cap S \cap T = \emptyset \text{ and } \forall \rho \in \mathcal{P}(N - (S \cup T \cup C)) : \\ v(C; \{S \cup T, C\} \cup \rho) < v(C; \{S, T, C\} \cup \rho)\end{gathered} \tag{3.28}$$

Analogously, we can define superadditivity and convexity for PFGs:

Definition 3.31. *A PFG is superadditive iff:*

$$\begin{gathered}\forall S, T \subseteq N \mid S \cap T = \emptyset,\ \forall \rho \in N - (S \cup T)\,, \\ v(S \cup T; \{S \cup T\} \cup \rho) \geq v(S; \{S, T\} \cup \rho) + v(T; \{S, T\} \cup \rho)\end{gathered} \tag{3.29}$$

Definition 3.32. *A PFG is convex iff:*

$$\begin{gathered}\forall S, T \subseteq N, \forall \rho \in \mathcal{P}(N - (S \cup T)) \\ v_\rho(S \cup T; \{S \cup T\}) + v_\rho(S \cap T; \{S \cap T, S - T, T - S\}) \geq v_\rho(S; \{S, T - S\}) + v_\rho(T; \{T, S - T\})\end{gathered} \tag{3.30}$$

A weaker concept than superadditivity employed in this dissertation is:

Definition 3.33. *A PFG is partially superadditive iff:*

$$\begin{aligned} &\forall P = \{S_1, \dots, S_m\} \in \mathcal{P}, \mid |S_i| > 1 \ \forall i = 1, \dots, k; \ |S_j| \ \forall j = k+1, \dots, m \ k \le m, \\ &\sum_i^k v(S_i, P) \le v(S, P') \ P' = P \backslash \{S_1, \dots, S_k\} \cup \{\cup_{i=1}^k S_i\} \end{aligned} \tag{3.31}$$

Partial superadditivity is weaker than superadditivity since it only non-singleton partitions.

Defining stability concepts for PFGs is more complex than for CFGs since now the value of a coalition depends on what other coalitions do. Economists have defined several core concepts based on design assumption of what other agents do when a particular coalition forms.

Definition 3.34. *γ-core:*

$$(x_1, \dots, x_n) \in \Re^n \ \mid \ \sum_{i \in S} x_i \ge v(S; S \cup [N - S]) \ \ \forall S \subset N \tag{3.32}$$

Here, $[N - S]$ represents the all-singleton partition formed by the agents belonging to $N - S$. The γ-core assumes that players in the deviating coalition believe that the rest of players will form singleton coalitions after their deviation. This concept is also called core with singleton expectations or s-core (Hafalir, 2007).

Another concept, the δ-core, assumes the total opposite: the belief is that the rest of players form only one coalition:

Definition 3.35. *δ-core:*

$$(x_1,\ldots,x_n) \in \Re^n \mid \sum_{i \in S} x_i \geq v(S; S \cup \{N - S\}), \ \forall S \subset N \tag{3.33}$$

Chander (2014) defines a concept that does not make such strong assumptions on the behavior of complementary agents:

Definition 3.36. *Strong-core (Chander, 2014):*

$$\begin{gathered}(x_1,\ldots,x_n) \in \Re^n \mid \forall P \in \mathcal{P}, P = \{S_1,\ldots,S_p\} \neq [N], \\ \exists S_i \in P, |S_i| > 1 \mid \sum_{j \in S_i} x_j \geq v(S_i, P) \ and \ if \ P = [N], \ x_i \geq v(i;[N]) \ \forall i \in N\end{gathered} \tag{3.34}$$

This concept of a core, used in section 5.1 is based on the notion of partial superadditivity, and it is related to the γ and δ cores. A payoff vector belongs to the strong-core if in every non-trivial partition (that is, one that forms from a deviation from the grand coalition) at least one non-singleton coalition is worse off, and in the trivial partition all coalitions (which all of them are singletons) are worse-off. This concept is behaviorally more interesting than the γ and δ-cores since it creates a coalition formation dynamics where the deviations from the grand coalition always end up forming the grand coalition back again. In other words, any deviation gains over the partition graph are unstable.

We now define the reduced game $< N, w^\gamma >$. In this game, the only partitions that can form are $P = \{S, [N - S]\}, \ \forall S \in 2^N$. Equivalently, we define the reduced game $< N, w^\delta >$ where the only partitions $P = \{S, N - S\} \ \forall S \in 2^N . w^\gamma(S)$ is the corresponding valuation function for the reduced game $< N, w^\gamma >$: $w^\gamma(S) = v(S, S, [N \backslash S]), \ \forall S \subset N$. Equivalently, $w^\delta(S)$ is the corresponding valuation function for this game: $< N, w^\delta >$: $w^\delta(S) = v(S, \{S, \{N \backslash S\}\}), \ \forall S \subset N$. The strong-core is non-empty under the following conditions:

Theorem 3.8. *(Chander, 2014)* $< N, v >$ *has non-empty strong-core if it has negative (resp. positive) externalities and the induced characteristic function game* $< N, w^{\gamma} >$ *(resp.* $< N, w^{\delta} >$*) is balanced.*

Theorem 3.9. *(Chander, 2014)* $< N, v >$ *with negative (resp. positive) externalities admits a non-empty strong-core only if the induced characteristic function game* $< N, w^{\delta} >$ *(resp.* $< N, w^{\gamma} >$*) is balanced.*

Theorem 3.10. *(Chander, 2014) If* $< N, v >$ *is partially superadditive, then strong-core =* γ*-core.*

Theorem 3.11. *(Chander, 2014)* $< N, v >$ *has a non-empty strong-core if it is partially superadditive and* $< N, w^{\gamma} >$*is balanced.*

With the concepts having been explained in this chapter, we can now proceed to their use in the analysis of five specific application contexts described in the following four chapters.

Chapter 4

Priced Exchanges in Traffic Intersection Control (PEXIC)

4.1 Introduction

Traffic intersection control has traditionally worked on a First Come First Served (FCFS) basis, since it is considered the fairest approach when treating all users anonymously and equally. Nevertheless, it is known that users do not value their travel times equally, since there exists heterogeneity in time, schedule and urgency costs. This conflict has been accepted and has remained as status quo for many decades. For instance, when discussing the objective function to be used for traffic signal optimization (Newell, 1989) assured that "[...] there is no mechanism whereby drivers can be charged for each service they receive. Even if there were, it probably would not be socially acceptable." While this statement has been true so far, it can be argued that developments in technology have brought about mechanisms that would allow users to be charged differentially for the services they receive. The focus in this chapter is in demonstrating such a mechanism.

This chapter proposes an exchange-based traffic signal control mechanism that incorporates the travelers' delay values in a scheme that increases intersection efficiency in terms of reduced travel costs. We show that treating users anonymously and reducing delay at intersections, as done right now, leads to unfair outcomes and economic inefficiency. We propose a control policy that is in fact Pareto-optimal and therefore Pareto-improving over the current schemes with delay minimization as well. Finally, the control mechanism can perhaps bring about social acceptability through individuals' acceptance, which in turn comes from reduction of unfairness with respect to others.

Since the control system no longer treats vehicles anonymously, travelers who would experience longer delays need to be compensated somehow by the travelers who experience shorter delays. This seeming unfairness is addressed by transferring utility from the former to the latter, in the form of monetary payments or equivalently, credit transfers. We are interested in achieving the fairest possible budget-balanced payments that support the most efficient allocation. These transfers result from a linear optimization program that minimizes how much users envy other users' supply allocations, in terms of delay and price. This approach falls within the framework of the compensation problem (Fleurbaey, 2008) found in normative welfare economics that addresses social acceptance of public policy.

In addition, we introduce for the first time the value of urgency of drivers into signalized intersection control. Recent empirical findings on High-Occupancy Tool (HOT) lanes (Bento et al., 2015) suggest that when users are given the opportunity, they pay mostly to not be late, not just to reduce their travel time. Operationally, the proposed system is designed to request users to input their valuation parameters and destination at the beginning of every trip. Users are not allowed to change it during the travel, or they are at least not allowed to change it frequently. This is designed to reduce incentives to misreport their urgency and to prevent individuals from selfishly manipulating the system outcome as they observe successive current states in the network.

Users will indicate their preferences by means of an app or an in-vehicle device. Valuations could be chosen from preset system recommendations or from a list of activities previously introduced in the app, to reduce user effort. The apps could also be learning programs that get fine-tuned based on the users' behavior. Various forms of Stated Preference (SP) surveys are also possible to elicit user-valuations. Users could also freely set that value if they wish. At the end of every trip, users would receive summary statistics on their experience and can be given the opportunity to reevaluate their preferences as a function of the realized payments and delay savings. It suffices to say that myriad mechanisms are possible for finding the valuations, and existing technology already allows it, though the schemes themselves remain to be developed.

We present a microeconomic analysis of an isolated signalized intersection when the travelers are allowed to exchange utility among themselves, with the goal of reducing travel cost and increasing fairness. Incorporating utility exchanges in traffic intersection control requires new multi-agent control algorithms since current schemes treat users anonymously and are based on average flows, rather than individual arrivals. We present the new concepts using optimization formulations that are in line with a state of the art in traffic control algorithms.

We analyze the asymptotic equilibria corresponding to long-run averages as a function of the input parameters. We introduce a novel pricing approach in transportation based on the concept of relaxed envy-freeness. To our knowledge this is the first attempt to use this concept in the transportation literature. In contrast to related previous research, our agents' value of delay savings is drawn from semi-empirical distributions that include the latest econometric results on urgency cost and scheduling delay.

The chapter is organized as follows: section 2 provides a brief description of the literature related to this chapter, section 3 introduces the method used, section 4 presents the phasing allocation algorithm, section 5 presents the pricing program and section 6 expounds the results, which are followed by a conclusion.

4.2 Literature review

The literature on intersection signal control is vast and well-established, and an elaborate review does not appear warranted here. Rather, the focus in this section is on the literature dealing with newer schemes that incorporate some form of user heterogeneity within the control. For instance, there has been recent interest in including priority at signalized intersections by using auctions. Most of these articles, however, use the Autonomous Intersection Management system (AIM) (Dresner and Stone, 2004) designed to control an intersection under a self-driving vehicle environment. Much of this recent literature is on techniques that may apply more to an autonomously controlled environment rather than one where each individual's behavior itself is considered. However, several concepts in this area of research have significance to the topic of this chapter, though it does not presume an autonomous environment.

(Schepperle and Böhm, 2007, 2008) introduce the concept of valuation traffic control, in which vehicles can reserve a slot in time-space bidding based on their value of time, and the control exploits this heterogeneity. They also introduce the concept of "subsidies" between vehicles, to facilitate high valuation vehicles behind low valuation vehicles to bid and get a reservation earlier. Unlike the approach we present in this chapter, these subsidies do not aim to compensate lower valuation vehicles for waiting more, but only to make high valuation vehicles of the same approach be considered in the control mechanism earlier. Delay reductions are observed in comparison to using FCFS, but the distribution of the values for time savings is arbitrarily set as exponential, which does not match with any empirical results and highly overstates efficiency gains due to its skewness to the right.

Carlino et al. (2013) uses second-price type auctions to establish an order to cross intersections, in which the sum of bids of the second-winning intersection approach is the payment from the winning approach, whose vehicles share the payment proportionally to their bids. Drivers never pay more than what they bid, but the mechanism does not ensure truthfulness. Drivers are assumed to bid using an automated trader called a "wallet". This policy with different bidding strategies (wallets) are simulated on real world city networks. Results are inconclusive since they depend on the network used, although the policy is clearly observed to increase efficiency, compared to the use of FCFS.

Levin and Boyles (2015) applies first-price auctions in isolated AIM controlled intersections to increase efficiency in the dynamic traffic assignment problem. Their article evaluates different auction policies, without or with subsidies, like in Schepperle and Böhm (2008). They found average travel time reductions for high value-of-time (VOT) vehicles and observed that longer queues can neutralize the policy benefits for high VOT vehicles. User valuations are drawn from a Dagum distribution based on the 2012 US Personal Income distribution. Even if value of travel time savings is tightly related to income, it increases less than proportionally. Directly using the distribution can potentially overstate efficiency gains due to a fatter right tail.

Vasirani and Ossowski (2008, 2010) extend the AIM to multiple intersections in a network. They recognize a traffic network to be a computational economy with two levels. At the lower level, intersections are considered agents which allocate time-space slots by means of combinatorial auctions. At an upper level, a general market equilibrium based pricing scheme controls for efficiency. Travel times are improved for those vehicles that are willing to pay but eventually benefits are neutralized as incoming flow increases, since average delay drastically increases, leading to a negative impact on social welfare.

Their valuation distributions are totally arbitrary.

In a later paper, Vasirani and Ossowski (2012) implement a combinatorial auction in single intersection control and examine the impact that a traffic assignment policy based on competitive markets on drivers' route choices. Finally, they consider both intersection control and traffic assignment policy on an adaptive management mechanism. They found that the distribution of vehicles on the network was more homogeneous leading to lower average travel times. Drivers with higher valuations got lower delays.

In conclusion, none of the above studies use empirical distributions of value of time savings or distributions of schedule delay savings, which are more relevant to the case of human drivers making decisions. Furthermore, most of them they assume exogenously defined budgets or distribution of bids with questionable economic foundation. Our research shows the high dependence of a control policy on the heterogeneity in the users' value of time savings and highlights the necessity of properly estimating this distribution. Moreover, all the cited articles use the reservation system (AIM) which requires a self-driving environment. Our approach is close to current practice, since it can operate within any look-ahead adaptive control via weighting of individual delays by individual values of delay savings.

The literature on adaptive traffic signal control system with conventional drivers is vast. Among the early systems were the well-known SCOOT (Robertson and Bretherton, 1974) and SCATS (Lowrie, 1990), which adapted the cycle length as a function of the demand volume after each period of time, normally around 5 minutes. In our case, we require adaptive control that is much more demand-responsive, which implies cycle-free algorithms. Some examples of this type of algorithms, typically solved through dynamic programming, are OPAC (Gartner, 1983), PRODYN (Henry, J.J., Farges, J.L., Tuffal, 1984), and the approaches in Kim et al. (2005) and Park (2009).

Our phase allocation control uses an adaptive, cycle-free, look-ahead algorithm that we develop for individual vehicle arrivals. This algorithm is in principle similar to the COP-type algorithms (Feng et al., 2015; Sen and Head, 1997) that are present in RHODES (Mirchandani and Head, 2001). A similar cycle-free algorithm can be found in Park (2009), which was also used for network-wide control in Yang and Jayakrishnan (2015). We chose this family of algorithms thanks to their polynomial computational complexity, but most of the adaptive control algorithms mentioned above could have been used as a starting point as well, with appropriate re-formulation for individual vehicle arrivals.

The work presented in this chapter conceptually views priced and brokered phase allocations in intersection control as a real-time exchange of supply at a signalized intersection, and establishes it with a microeconomic basis. Such a supply-exchange perspective in signal control is new, to our knowledge. Our objective is to increase efficiency at intersections by reducing travel costs while addressing fairness for public acceptance. Individuals' values of delays are based on the most recent and relevant empirical value of time savings studies.

4.3 Overview of the Method

To see a simple illustration of the concept, consider the intersection example shown in Fig. 4.1. At time instant 0, the orange and blue vehicles are queued from the previous rolling horizon. Their delay values are θ_1 and θ_2. The time to clear the queue t_q is lost time L plus twice the saturation headway h plus the all-red period r. The yellow vehicle with delay value θ_3 will cross the intersection in t_a seconds. Suppose $t_a < t_q$. By minimizing the sum of total costs, the product of value by delay, it is easy to verify that the yellow vehicle will not stop as long as $\theta_3/(\theta_1+\theta_2) > (t_a+r)/(2L+3h+r-t_a)$. This is true for instance for $\boldsymbol{\theta} = \{10, 20, 60\}$ \$/h and $L = 1, hL = t_a = 5, r = 2$. On a delay minimization setting the

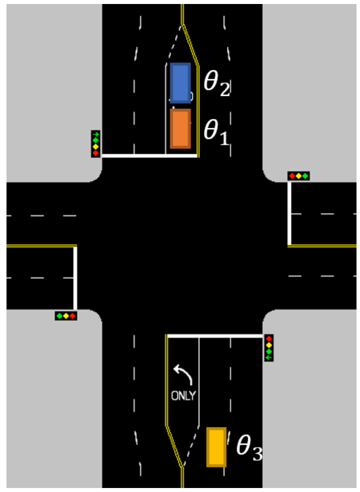


Figure 4.1: Intersection example

queued vehicles would be given green first, leading to a total cost of $0.20. In our cost minimization setting, the total cost would instead be lesser, at $0.18.

With these specific parameter values, we see that this control system increases economic efficiency. Since this control policy alters delays as a function of individual characteristics, we need to compensate the delay differences with payments, to make the policy fair with respect to the status quo, namely delay minimization. This requires monetary exchanges across the drivers (which can be brokered though an intersection manager or a centralized Traffic Management Center), and thus a pricing scheme is required. We name this general framework of traffic control as Priced Exchanges in Intersection Control (PEXIC).

However, notice that the above example required the yellow vehicle to have a sufficiently high value of delay, which implies that sufficient heterogeneity in values of delay is required for such economic efficiency to be achieved. The value of delay of any driver depends on a variety of variables such as the innate value of time (VOT) of the individual, as well as the perceived time pressure due to perceived lateness at the intersection, which in turn depends on the expected arrival time at the destination. Thus any delay saved by this control method has a value, which we will call Value of Delay Savings (VDS). In the following discussions, we provide simple and plausible specifications for VDS, a control algorithm that uses VDS, as well as an associated pricing scheme that compensates users appropriately, for an equitable and fair system.

In the following discussion, for simplicity we use the term "vehicle" to also refer to the groups of travelers in a vehicle, and assume that each vehicle represents a composite individual with a certain behavior with respect to delays, while we recognize that complex mechanisms to find the composite behavior of a group of travelers require studies which are beyond the scope of this first study of the concept. Each vehicle i is assumed to have a quasilinear utility function as next:

$$u_i = v_i - \pi_i = -\theta_i d_i - \pi_i \tag{4.1}$$

Where, for vehicle i : v_i = delay valuation (called the "valuation" of the vehicle), θ_i = the value of delay savings (VDS), (called also the "value" of the vehicle), d_i = control delay, π_i = the price charged.

We define the mechanism M as a pair of functions $k : \Theta \to K,\ p : \Theta \to \mathbb{R}$. We call the first function the phase allocation rule that maps the type space Θ to the allocation space K. The second function is the payment rule, which maps Θ to the real line, the payment space. In our notation k can represent either the allocation function or an element in the allocation space K. We assume complete information and therefore there are no strategic implementation issues.

The phasing allocation rule is first solved to determine the order and length of phases during a rolling horizon of length T seconds. It minimizes the sum of all valuations that are part of the rolling horizon. This is done using a formulation that includes complete 8-phase dual ring NEMA control with non-conflicting simultaneous movements and barrier groups, as described in section 4 below. A barrier group is a set of movements all of which originate from the same street (i.e. North-South Street or West-East Street in Fig.4.1) and are separated from the other barrier group by a barrier which sets all movements to red. The fair price program charges a price $p_{i,h}$ to each vehicle $i \in I_h$, where I_h is the set of vehicles present during rolling horizon h. Please note that these prices can be positive, when the vehicle pays money, or negative, when the vehicle receives money.

Once a full rolling horizon optimization for horizon h (phases and prices) is carried out, we let a rolling period of length equal to the length of the first barrier group, i.e. bg_1^h, pass and re-optimize for a new horizon. Referring to (4.1) above, π_i is the sum of all $p_{i,h}$ such that $I_h \ni i$, that is, the sum of every partial price vehicle i is charged in every rolling horizon h that i participates in, until departure from the intersection. Fig.4.2 shows this rolling horizon scheme.

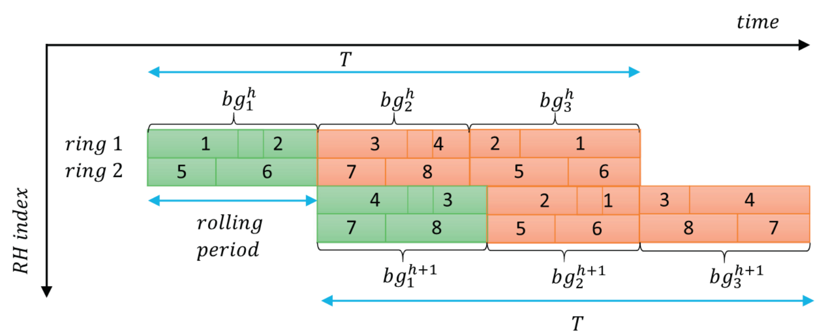


Figure 4.2: Rolling horizon scheme

Before proceeding to describe the optimization schemes and the envy-free pricing principles, we present the fundamental microeconomic concepts on which PEXIC's theoretical properties are based. These properties hold regardless of the type of signal control formulation and algorithm used for the allocation rule, and the payment rule used in the mechanism, as long as the quasilinear utility function assumption, found in (4.1) holds. Next we define our efficiency measures:

$$u_i(k', \theta_i) > u_i(k, \theta_i) \Rightarrow \exists j \in I \mid u_j(k', \theta_i) < u_j(k, \theta_i) \tag{4.2}$$

That is, changing the allocation only improves some agents' utility but at least one agent is worse off.

Definition: a mechanism M is allocative-efficient if for any type vector $\boldsymbol{\theta}$:

$$\sum_{i \in I} v_i\left(k(\boldsymbol{\theta}), \theta_i\right) \geq \sum_{i \in I} v_i\left(k', \theta_i\right) \quad \forall k' \in K \tag{4.3}$$

This means that there is no allocation k' such that the sum of individual valuations is larger than under the sum of individual valuations under k.

Definition: A mechanism M is budget balanced if and only if for any type vector $\boldsymbol{\theta}$:

$$\sum_{i \in I} p_i(\boldsymbol{\theta}) = 0 \tag{4.4}$$

Theorem 4.1. *A mechanism that is allocative-efficient and budget balanced under quasilinear preferences is Pareto optimal.*

Proof. See proof in Theorem 3.1 □

We thus show the PEXIC framework as allocative efficient, budget balanced and Pareto efficient (optimal) for every optimization step. These properties can hold under many types of signal control formulations and algorithms used as allocation rules and also many types of payments rules, such as the VCG mechanism (Clarke, 1971; Groves, 1973; Vickrey, 1961).

4.4 Phase allocation algorithm

Current adaptive intersection control algorithms treat users anonymously, and normally they are aggregated into flows or counts. Treating users individually requires keeping track of each individual, their characteristics and intermediate outcomes during the calculations. While the main general structure does not necessarily need to be modified, the basic data structures are no longer integer arrays but multi-attribute lists which are updated at every relevant calculation. These aspects significantly increase the computational complexity of the algorithm.

The phasing allocation algorithm can be summarized by the following optimization problem for a given horizon h of control. We have an objective function F which denotes the sum of total vehicle cost. Vehicles' costs arise from the optimal departure times, found in the departure time list V_h which obviously depends on the signal timing, expressed by the decision variable X. At the same time, the objective depends on the input arrival times and individual VDS values, represented by the input list I_h. The objective function is subject to signal timing constraints, represented by the constraint set Ω.

$$\min_X F(V_h(X), X; I_h) \tag{4.5}$$

$s.t.$

$$\Omega(V_h(X), X; I_h) = 0 \tag{4.6}$$

The conceptual optimization in (4.5) and (4.6) is fundamentally different from the traditional signal optimization schemes in which individual vehicle variables are not considered.

The function F is obviously not simple thanks to individual vehicle variables, unlike in traditional flow-based optimizations. In order to reach practical computational times for our adaptive, cycle-free and look-ahead phase allocation algorithm, we chose as a starting point the recursive structure of COP-type algorithms (Feng et al., 2015; Sen and Head, 1997) due to their polynomial complexity. In contrast with traditional formulations which represent queues as scalar values, our queues are lists of vehicles with updated multiple individual attributes to incorporate PEXIC's multi-agent nature. Thus every vehicle cost is considered individually and not aggregated by flow intervals. These multi-attribute lists need to be updated at every stage-state and the internal algorithm operations are no longer just additions and subtractions, but also list operations and update of attributes' values. Finally, the individual VDS parameters $\theta_i \in \Theta$ are introduced.

The algorithm is essentially based on Dynamic Programming and consists of a forward recursion which at every stage-state selects the best feasible control decision, i.e. the barrier group length with phase order and phase lengths. This is followed by a backward recursion to retrieve the optimal control policy. Each control decision performance function is evaluated by carrying out an exhaustive search over all possible green lengths and phase orders within a barrier group, for every control decision for every state. In the appendix A of this dissertation, a summary of the main variables used in PEXIC algorithm can be found.

Let T be the rolling horizon length. Let I_h be the list of vehicles that participate in rolling horizon h. This list comprises the approaching vehicles during h as well as the queued vehicles from the previous horizon. The time step granularity is 1 second in our implementation. $j \in 1 \ldots J$ is the index of every stage (barrier group). The number of stages J is determined once the stopping criterion is met. The two barrier groups $b \in \{0,1\}$, are iterated alternatively, i.e., $b(j) = 1 - j\text{mod}(2)$. For every stage j, we define the state variable s_j as the number of time steps between the start and the end of barrier group number

j. Therefore, potential feasible barrier groups will have length $x_j = s_j - s_{j-1}$. The control decisions x_j are constrained by the following minimum and maximum barrier lengths:

$$X_j^{min} = \max\{G_{min}^{1,1} + Y^{1,1} + G_{min}^{2,1} + Y^{2,1},\ G_{min}^{1,2} + Y^{1,2} + G_{min}^{2,2} + Y^{2,2}\} \tag{4.7}$$

$$X_j^{max} = \min\{G_{max}^{1,1} + Y^{1,1} + G_{max}^{2,1} + Y^{2,1}\ ,\ G_{max}^{1,2} + Y^{1,2} + G_{max}^{2,2} + Y^{2,2}\} \tag{4.8}$$

$G_{min}^{p,r}$ and $G_{max}^{p,r}$ are the minimum and maximum green times for phase p in ring r. And $Y^{p,r}$ is the yellow plus clearance time of phase p in ring r. $G_{min}^{p,r}$ are set to 5 seconds and $G_{max}^{p,r}$ to 50 seconds. $Y^{p,r}$ are set to 4 seconds. The control variables $x_j \in X_j(s_j)$ are defined from the previous states s_{j-1} as $z_j\left(s_{j-1}\right) \in Z_j\left(s_{j-1}\right)$, but they are implemented in practice from s_j as in the original COP algorithm. Thus:

$$Z_j\left(s_{j-1}\right) = \begin{cases} \{0\} & if\ \ T - s_{j-1} < X_j^{min} \\ \{0\} \cup \{X_j^{min}, \ldots,\ X_j^{max}\} & if\ \ X_j^{max} \leq T - s_{j-1} \\ \{0\} \cup \{X_j^{min}, \ldots, T - s_{j-1}\} & if\ \ X_j^{max} > T - s_{j-1} \end{cases} \tag{4.9}$$

$$X_j\left(s_j\right) = \left\{z \in Z_j\left(s_{j-1}\right) |\ s_{j-1} + y = s_j\right\} \tag{4.10}$$

Note that barrier groups can be skipped by allowing $x_j(s_j)$ to be zero.

For every stage $s_j \in S_j$ and feasible control decision $x_j\left(s_j\right) \in X_j\left(s_j\right)$, the forward recursion evaluates the performance index $f_j(s_j, x_j)$ and the resulting temporary vehicle lists $L_j^D(s_j, x_j)$ and $L_j^Q(s_j, x_j)$. The performance index is the sum of all the individual costs of all the vehicles that have arrived, departed or been queued between $s_j - x_j$ and s_j. Then it updates the value function $v_j(s_j)$ by selecting optimal control decision x_j^* and selects the corresponding two temporary optimal lists. $D_j\left(s_j\right) = L_j^D(s_j, x_j^*)$ contains all the vehicles that departed during $(s_j - x_j^*, s_j)$ and $Q_j\left(s_j\right) = L_j^Q(s_j, x_j^*)$ contains all the vehicles that were in queues at the end of that interval. The forward recursion ends when the stopping criterion is met. Essentially, the stopping criterion first makes sure that all barrier groups have been considered ($J > 2$) and then asserts that repeating them does not improve the value function.

During the backwards recursion the optimal plan s_j^* is retrieved and final vehicle list V_h is built by appending the successive departed vehicle lists of the optimal states s_j^* and the last queued vehicles list $Q_{J-1}(s_{J-1}^*)$ at state T. V_h contains the optimal vehicle departure times that will serve to compute the prices later. The pseudocode for the phasing algorithm is:

Each performance measure $f_j\left(s_j, x_j\right)$ is calculated with an exhaustive search over the set of feasible green times for each of the two rings and the four possible phase orderings in the barrier group $b(j)$. It is represented by the following optimization problem:

$$\min \sum_{r\in\{1,2\}} \sum_{c\in C} z_{r,c} \left(\sum_{p\in\{1,2\}} \sum_{i\in I_{h,p}} \theta_i \left(t_i\,(c,\ g_r) - \max\{a_{i,h}, s_{j-1}\}\ \right) \right) \tag{4.11}$$

$$s.t.$$

$$z_{r,c} \in \{0,1\}\ \forall r, c \in \{1,2\} \tag{4.12}$$

$$\sum_{c\in\{1,2\}} z_{r,c} = 1 \forall r \in \{1,2\} \tag{4.13}$$

$$\phi_{1,c} = 1 + z_{r,c}\phi_{2,c} = 2 - z_{r,c} \forall r, c \in \{1,2\} \tag{4.14}$$

$$\max\left\{G_{min}^{p_{1,c},r}, x_j - Y_{p_{1,c},r} - Y_{p_{2,c},r} - G_{max}^{p_{2,c},r}\right\} \leq g_{r,c} \forall r, c \in \{1,2\} \tag{4.15}$$

$$g_{r,c} \leq \min\left\{G_{max}^{p_{1,c},r}, x_j - Y_{p_{1,c},r} - Y_{p_{2,c},r} - G_{min}^{p_{2,c},r}\right\} \forall r, c \in \{1,2\} \tag{4.16}$$

$z_{r,c}$ are binary variables which take the value of one if for ring r, phase order c is selected and zero otherwise. (4.14) makes only one ordering per ring possible. (4.15) uses $z_{r,c}$ to

Algorithm 1 PEXIC Phase Allocation

procedure PHASE ALLOCATION
Input: I_h
Output:V_h
$j = 1,\ s_0 = 0, x_0 = 0,\ v_0 = 0$, stop = **FALSE**
while (stop == **FALSE**)
build $X_j\left(s_j\right)$
for s_j **in** S_j
for x_j **in** $X_j(s_j)$
$s_{j-1} = s_j - x_j$
$v_j'\left(s_j, x_j\right) = f_j\left(s_j, x_j\right) + v_{j-1}(s_{j-1})$
build $L_j^D(s_j, x_j)$
build $Q_j^D(s_j, x_j)$
endfor
$x_j^* = argmin_{x_j}\left[v_j'\left(s_j, x_j\right)\right]$
$v_j(s_j) = v_j^{'}\left(s_j, x_j^*\right)$
$D_j\left(s_j\right) = L_j^D(s_j, x_j^*)$
$Q_j\left(s_j\right) = L_j^Q(s_j, x_j^*)$
endfor
if (j≤2)
j = j+1
stop = **FALSE**
elseif ($v_{j-k}(T)$==$v_j(T)$) **forall** k≤ 1
break
else
j = j+1
stop = **FALSE**
end
endwhile
$s_{J-1}^* = T$
$M_{J-1} = Q_{J-1}(s_{J-1}^*)$
for $j = J-1$ **to** 1
read $x_j^*(s_j^*)$
$s_{j-1}^* = s_j^* - x_j^*(s_j^*)$
$V_h = D_{j-1}\left(s_{j-1}^*\right) \frown M_j(s_j^*)$
endfor

select the actual phase orders defined by phase indices $\phi_{1,c}$ and $\phi_{2,c}$. Vehicle departure times t_i are calculated with a vertical queue logic, **Algorithms 2a** and **2b**, relative to the rolling horizon start. (4.16) defines the set of feasible green times $g_{r,c}$. t^{endq} are the departure times of the last queued vehicle of every phase. The vertical queue logic uses accumulated saturation headways $h_{sat}(i)$ that could be calibrated on site (or from the simulation, as done in the studies in this chapter) using observed queue departures.

Algorithm 2a: queue logic 1st phase in ring r	**Algorithm 2b**: q.l. 2nd phase in ring r
Input: $I_{h,1}$	**Input**: $I_{h,2}$
Output: $\boldsymbol{t}$	**Output**: $\boldsymbol{t}$
$t^{endq} = s_{j-1}$	$t^{endq} = s_{j-1} + g_{r,c} + R_{p_{1,c},r}$
for $i \in I_{h,1}$	**for** $i \in I_{h,2}$
if $(g_{r,c} > h_{sat}(i)) \wedge \left(a_{i,h} < t^{endq}\right)$	**if** $\left(x_j - R_{p_{2,c},r} + s_{j-1}\right) \wedge \left(a_{i,h} < t^{endq}\right)$
$t_i = L + (i-1)\, h_{sat} + s_{j-1}$	$t_i = h_{sat}(i) + s_{j-1} + g_{r,c} + R_{p_{1,c},r}$
$t^{endq} = t_i$	$t^{endq} = t_i$
elseif $\left(a_{i,h} > t^{endq}\right) \wedge \left(g_r + s_{j-1} > a_{i,h}\right)$	**elseif** $\left(a_{i,h} > t^{endq}\right) \wedge \left(s_{j-1} + x_j - R_{p_{2,c},r} > a_{i,h}\right)$
$t_i = a_{i,h}$	$t_i = a_{i,h}$
else	**else**
$t_i = s_{j-1} + x_j$	$t_i = s_{j-1} + x_j$
endif	**endif**
endfor	**endfor**

The first COP algorithm (Sen and Head, 1997) is known to be $O(T^3)$(Shelby, 2004; Samra et al., 2011). We show next how PEXIC's multi-agent nature alters this complexity. Let T be the maximum number of time steps per stage. For every stage j, the feasible decision set $X_j(s_j, x_j)$ has a maximum size of $\Delta X^{max} = \Theta(\max\left(X_1^{max} - X_1^{min}, X_2^{max} - X_2^{min}\right) + 1)$ and there is a maximum of T states per stage. We now need to calculate the maximum number of stages. Assuming a worst case scenario in which every stage has the minimum length $X^{min} = \min(X_1^{min}, X_2^{min})$, the maximum number of stages is $O\left(T/X^{min}\right) < O(T)$. Calculating $f_j(s_j, x_j)$ takes place then in $O(TN)$ time, where N is the number of vehicles participating in the rolling horizon and it can be expressed as the product of the total average flow q and T. The final computational complexity of the allocation rule is thus $O\left(T^4 N\right) = O(T^5 q)$.

The pruning method employed in the forward recursion is what makes the allocation rule an approximate algorithm for the phasing, but it is crucial to achieve real-time solutions. The loss in efficiency of the original COP due to the pruning was empirically bounded to 6% by (Shelby, 2001). Even though we did not test the efficiency loss in this dual-ring version, we expect it to be smaller, since the stages are now full barrier groups instead of phases, reducing the final number of stages in the optimization. Nevertheless, using a rolling horizon approach is itself a source of inefficiency and the pruning is thus only a minor issue.

Rolling horizon length T is defined exogenously in our model. On the one hand, we would desire T to be as long as possible to include as much as information from future vehicles and increase the quality of the prediction. However, we do know that the quality of the information decreases as we look further into the future. Some authors attempted to apply decreasing weights as a function of time on the delays or incoming flows to give more importance to earlier arrivals. These functional specifications did not generally give better results than no weighting at all. Finally, a longer rolling horizon is computationally more expensive. In the present isolated intersection setting, T should be at least larger than the maximum length of the first barrier group (under maximum expected flow) plus the minimum length of a second barrier group. Thus, we set here T to 80 s.

An alternative discussed in (Newell, 1998) is to include a salvage cost, that is, a value that accounts for value or cost that is not considered in the current rolling horizon itself.

He concluded that including a salvage cost is equivalent to extending the rolling horizon length. In our particular case of multi-agent control, such a salvage cost expression would be very complex, since it depends on each of the individual value of delay savings. Such a salvage cost expression should first account for the cost incurred during further rolling horizons by vehicles present in the current rolling horizon h which do not depart in the current rolling horizon. Second, it needs to account for the cost caused by vehicles who have not departed in h on the vehicles who are not present in the current rolling horizon. Preliminary studies showed that none of the attempted terminal cost expressions showed any improvement at all, over the results from simpler scheme of using no salvage cost and no weighting.

As would be clear to the reader by now, the value of delay savings of individuals is intricately tied to the above phasing algorithm. The question arises as to the kind of values to be used in optimization. The next subsection provides the pointers on an appropriate treatment of this variable based currently available empirical results.

4.4.1 Value of delay savings (VDS)

The most influential parameter in this control is naturally the value of delay savings (VDS) θ. Thus a detailed discussion is warranted, before it is incorporated in the signal phasing algorithm. As these are not observed at intersections and no data is available, the initial studies require assumptions on what distribution over what range of values to assume for VDS, across the driver population. For the purposes of this chapter, we assume that for each user i, it is appropriate to draw individual θ_i values from a semi-empirical distribution, on the basis of the latest empirical findings on drivers' value of travel time savings, scheduling costs and value of urgency. This is in line with two empirical studies that appear to be relevant. Abou-Zeid et al. (2010) obtained a VOT distribution estimated from a Swedish SP survey based on car alternatives, accounting for income levels and latent variables. This VOT distribution serves as a base layer for our assumed value of delay savings (VDS) distribution. Recent evidence from High-Occupancy Toll lanes suggest though, that drivers are not willing to pay just to reduce the time spent on the road, but that they may pay to avoid arriving late or at least to reduce the lateness. This leads to models with more parameters than just the value of time.

There are dynamic models with endogenous scheduling that include more parameters in vehicles' utility function (Arnott et al., 1990; Small, 1982). Vehicles are assumed to choose a departure time t and are willing to arrive at the destinations at a desired arrival time t^* after having crossed the bottleneck. Vehicles suffer late or early schedule delay costs if their arrival time $t + T(t)$ is larger or smaller than t^*, respectively, where $T(t)$ is the travel time for departing at t. A recent study (Bento et al., 2015) adds a fourth parameter, urgency cost δ, which is a constant term for being late. This extension is capable of explaining observed willingness to pay per hour for small travel time savings. The travel cost expression c becomes:

$$\begin{aligned} c_i(t|t^*, \alpha, \beta, \gamma, \delta) = \alpha \cdot T(t) + \beta \cdot (t^* - t - T(t) | t + T(t) \leq t^*) + \\ \gamma \cdot (t + T(t) - t^* | t + T(t) > t^*) + \delta \cdot I(t + T(t) > t^*) \end{aligned} \tag{4.17}$$

This model requires a desired arrival time for every driver. Unfortunately, no data is available on the desired arrival times of people at any given time in a traffic network. Bento et al. (2015) estimates that around 7% of the drivers in their data would be running late at any time, in freeway contexts. In their bottleneck models without urgency costs this percentage increases to 20%. Their definitions of lateness, however, may not apply to what fraction of drivers perceive that they are late at any intersection.

Outside of academic economics, surveys in the domain of human resources have studied lateness in the workplace in North America and UK (Careerbuilder.ca, 2015; Tiptopjob.com, 2004; Yougov.us, 2014). By aggregating the lateness frequency categories for every study, we obtain the percent of workers being late or the probability of one random worker being late, x. These final probabilities oscillate between 10% and 23%. For the simulation experiments here, we study varying this parameter roughly in this range. Since the evidence on lateness is based only on one activity, commuting for work, testing other values allows us to consider other trip activities in which lateness may be less prevalent; e. g. groceries purchase.

The VDS distribution used is determined as follows. As a first step, the value of time α_i is drawn from Abou-Zeid et al. (2010). Secondly, each driver is assigned the status of currently expecting to be late or not, using a simple Bernoulli trial with a success-probability of x_i. If the driver is late, a late schedule delay coefficient γ is added (x_i is set to one). According to Bento et al. (2015), the lateness-cost, γ, is two times larger than the earliness cost, α. In the present analysis, the cost of earliness is anyway not considered. Thus, our values of delay savings will be drawn from a mixture distribution in which some vehicles are late and some are not.

Separately, an urgency cost is introduced following Bento et al. (2015). δ_i is divided by a fixed amount of minutes m the vehicle is expected to be late. The urgency felt at each intersection by any given driver may depend on where on the driver's trip route the intersection is. This is somewhat different than in Bento et al. (2015), where m would be equal to the time savings of using the express lane in comparison to using the main lanes. Since the actual time savings per intersection are very small and the traffic conditions (local queues) are more volatile than those of an express lane with limited entry and exit points, m is assumed to be the minutes that vehicle is late for a particular corridor.

Given that there is no data available about how late people are in an urban network, we assume that the lateness (in minutes) follow an exponential distribution shifted by five minutes and with a parameter λ equal to $1/5\ \text{min}^{-1}$ to achieve an average lateness of 10 minutes. That is, $m \sim 5 + \exp(\lambda)$. The shifting parameter serves both as a lower threshold of lateness, as actually observed in questions about lateness valuation from employers (Yougov.us, 2014), and also to avoid abnormally high values that would have excessive influence during the phasing. Thus, the VDS θ_i is:

$$\theta_i = \alpha_i + x_i \left(\gamma_i + \frac{\delta_i}{m} \right) \tag{4.18}$$

This is a simple model, though it has some empirical justification. We use this model in the control schemes, the details of which are presented in the next section.

The above model could be equivalently used to incorporate carpool vehicles. A within-vehicle value-aggregation scheme would provide a joint value for each vehicle and indirectly handle carpooling and urban logistics vehicles. This would add extra skewness in the VDS distribution which can potentially yield even higher efficiency levels. Such possibilities are however not considered here.

Note also that neither vehicle-type heterogeneity nor pedestrians are considered here. A similar value-aggregation scheme, this time considering different queuing space and different accelerations could be used to account for buses and urban logistics vehicles, which are slower than cars and use more space in the queue. In addition, pedestrians could be incorporated by detecting their position and speed through wireless communication or with the pedestrians announcing their presence using NFC or RFID, in a similar way as they currently use pedestrian-buttons.

From an environmental or social point of view, pedestrians are preferred to vehicles since they do not pollute or create congestion. Alternatively, public administration could define an appropriate average value for pedestrians to favor pedestrian modes in particular areas. Pedestrian arbitrage could be prevented by identifying abnormal or repetitive crossing patterns. Alternatively, pedestrian heterogeneity could be entirely dropped, each pedestrian being assigned a predefined value.

4.4.2 Envy-minimizing Pricing scheme for Fairness

The purpose of the price rule is to compensate the differences in delays caused by the inclusion of VDS heterogeneity in the phase allocation algorithm and not treating users anonymously. These differences need to be addressed as it brings up a public acceptance issue, and also to guarantee user satisfaction. We address this problem from a normative economics point of view, by maximizing fairness across individuals. As fairness is difficult to objectively define, we address it by means of the most common approach found in literature, namely envy-freeness. An allocation is envy-free if there is no pair of users that would like to exchange their allocations at the current prices. Alternatively, no agent values anybody else's allocation at the given price more than their own. This is translated in our problem by the following conditions:

$$-\theta_i\left(t_i - a_i\right) - p_{i,h} - \pi_{i,h-1} \geq -\theta_i\left(t_j - a_j\right) - p_{j,h} - \pi_{j,h-1} \qquad \forall i, j \in I_h, \quad i \neq j \tag{4.19}$$

For a total number of N_h participants in a rolling horizon, there will be $N_h(N_h-1)$ of such constraints. Where $\pi_{i,h-1}$is the accumulated price from previous rolling horizon optimizations. Each arrival a_i is the first projected arrival since vehicle participates in the mechanism.

Users suffer a particular degree of envy under a particular state defined by a vector of individuals and their associated delays and prices. These envy values only changes once the state changes. Since we consider that a predicted arrival time does not change after the vehicle is detected, the only two elements that can alter envy levels are changes in prices and departure times. This only happens once new vehicles are detected or some present vehicles depart. This means that if a particular state lasts several seconds, the envy levels should remain invariant.

Specifying the envy-free conditions as a function of accumulated predicted delays and accumulated prices makes the envy-free conditions independent from the frequency of optimization. Thus, if we wished to optimize the phasing more frequently and if there were no changes in the participant set during a particular period, the envy levels would remain invariant. Furthermore, it also prevents double counting in prices, despite the rolling horizons being overlapped. Finally, note that the utility comparisons done by any agent do not use the utility specifications of the other agents being compared against. That is, each agent only uses its own utility specification for the valuation of the allocations for other agents. Thus envy-freeness has low informational requirements.

Unfortunately, completely envy-free allocations may not exist. A necessary condition for an allocation to be envy-free is swap-monotonicity in the allocation (Hartline, 2014). An allocation is swap-monotone if and only of after sorting participants' values in decreasing (increasing) order, the allocation amounts are decreasing (increasing). In our application this translates as $\theta_1 \geq \cdots \geq \theta_n \implies d_1 \leq \cdots \leq d_n$. This condition rarely holds for a traffic intersection, especially as the number of incoming vehicles increases. For instance, departing vehicles on one approach may not necessarily be ordered by valuation, but the delays are ordered increasingly with the queue position. This is the case of the introductory example shown in Fig. 4.1.

Unless the efficient allocation obtained from the allocation algorithm is swap-monotonous, envy-free prices that support it will not exist. A way to circumvent this problem is to relax the envy free conditions and allow agents to suffer a degree of envy. In this case, the payments seek to support the efficient delay allocation causing as low a level of envy as possible. We relax these conditions by adding a slack term $\epsilon_{ij} \geq 0$ which represents the minimum amount of excess envy agent i suffers on comparisons against every other agent j. Thus:

$$\epsilon_{ij} - \theta_i (t_i - a_i) - p_{i,h} - \pi_{i,h-1} \geq -\theta_i \left(t_j - a_j\right) - p_{j,h} - \pi_{j,h-1} \quad \forall i,j \in I_h, \quad i \neq j \tag{4.20}$$

There is an extensive body of literature in welfare economics and social choice theory on relaxing and refining envy-freeness. Most of these articles (Chaudhuri, 1986; Diamantaras and Thomson, 1990) address the envy minimization problem within an exchange economy framework. In these articles, both allocation and prices are optimized simultaneously. In our case, delays and prices are optimized sequentially. The prices serve to compensate for the differences in delays imposed by the previous step. This kind of problem is called "compensation problem" in welfare economics literature (Fleurbaey, 2008).

Compensation problems aim to compensate for inequalities in individual characteristics by means of transferrable resources, i.e. monetary payments or credit transfers. In compensation problems, individuals are considered to be endowed with two types of characteristics: the ones for which they are considered not responsible (i.e. circumstances) and the ones for which they are. We consider the delays d_i as a "not-responsible" characteristic, since they are previously fixed by the phase allocation algorithm and individuals do not have a direct say in them. The VDS θ_i are then a "responsible" characteristic, since they are directly chosen and input to the system by the individuals themselves.

Let E be the domain of delay-value situations $e = (\boldsymbol{d}, \boldsymbol{\theta})$. Let P be the set of feasible budget balanced price allocations:

$$P = \left\{ \boldsymbol{p} \in \mathbb{R}^{N_h} \mid \sum_{i \in I_h} p_i = 0 \right\} \tag{4.21}$$

An allocation rule is a correspondence $f : D \rightarrow\rightarrow P \mid \forall e \in E$. Then, $f(e) \subset P$ is the subset of allocations selected by S. Allocation rules found in literature are mostly minimax based, see Fleurbaey (2008). In our preliminary studies, such minimax approaches were found not to be sufficiently redistributive, in the sense that low value vehicles' average utility was not higher than in the case of delay minimization. In other words, the policy was not Pareto-improving in comparison to delay minimization. This is because such minimax criteria disregard envy terms that are not binding at the optimum. For this reason, we used a more flexible ad-hoc approach that helps our control algorithm to be Pareto efficient in comparison to delay minimization. This approach consists of using a weighted additive objective function composed of the envy terms ϵ_{ij} and weights $\gamma_i \in (0,1)$.

Definition 4.1. *(weighted) additive envy allocation rule:*

$$\forall e \in E,\ \forall p \in P:\ p \in f(e) \Longleftrightarrow \forall p' \in P, \sum_{i,j \in I_h \mid i \neq j} \gamma_i \epsilon_{ij}(p) \leq \sum_{i,j \in I_h \mid i \neq j} \gamma_i \epsilon_{ij}(p') \tag{4.22}$$

In other words, there is no price vector p' such that the objective function value under that vector is lower than in the price vector p set by the allocation rule. Additive criteria are an easy way to incorporate all relations since all the individual envy terms behave as substitutes. Obviously this departs from the common notion of envy in which an agent's envy is the maximum of the individual envies. However, the presence of complex allocative externalities across vehicles makes vehicles' delay depend on each other. Our context of intersection control, wherein a group of vehicles on a conflicting movement are given priority over each vehicle in a given (stopped) movement, is different from the relatively simple cases of resource allocation considered in theoretical economics literature, and thus a joint weighted additive form may not be unrealistic to assume. Nonetheless, individual envies will be examined in terms of both maximal and additive metrics here.

We assume further that the weights γ_i follow a logistic specification form leading to monotonic decreasing weights with respect to VDS. This functional form makes the lower VDS-associated envy terms more costly than the high ones. Consequently, envy terms of the low value vehicles are minimized further than those of the high value vehicles, leading to higher average payments for the latter. The logistic specification was chosen for its flexibility and simplicity, and also for its suitable bounds. The actual weight expressions are:

$$\gamma_i = \frac{1}{1 + ae^{-\frac{b(\overline{\theta}_h - \theta_i)}{\overline{\theta}_h - \underline{\theta}_h}}} \qquad \forall i \in I_h \tag{4.23}$$

The exponent is an affine linear function of the basic statistics, minimum and maximum vehicle values among all the vehicles present in horizon h. This implies that we do not assume knowledge of the population distribution. The functional form assumed for γ_i involves two parameters a and b, a base intercept and the steepness of the curve respectively, which can be learnt on the field. The purpose of this weight calibration by exhaustive enumeration is to show that the proposed mechanism is also Pareto-improving (efficient) with respect to delay minimization.

Once a specification is selected for envy minimization criteria, we can proceed to the optimization that results in the pricing needed as the compensatory mechanism. The pricing scheme is then based on the following linear pricing program (LP):

$$\min \sum_{i,j \in I_h | i \neq j} \gamma_i \epsilon_{ij}(p) \tag{4.24}$$

$$s.t.$$

$$\epsilon_{ij} - \theta_i d_i - p_{i,h} - \pi_{i,h-1} \geq -\theta_i d_j - p_{j,h} - \pi_{j,h-1} \qquad \forall i,j \in I_h, \quad i \neq j \tag{4.25}$$

$$\sum_{i \in I_h} p_{i,h} = 0 \tag{4.26}$$

$$\epsilon_{ij} \geq 0 \qquad \forall i,j \in I_h, \quad i \neq j \tag{4.27}$$

Here the objective function (4.24) is a weighted additive criterion and the constraints (4.25), (4.26) and (4.27) refer to the relaxed envy-free conditions, budget balancing, and the lower bounds for the envy terms to make them positive. We can prove though, that whenever the set of envy-free allocations is non-empty, our social ordering function will select it.

Proposition 4.1. *The weighted additive allocation rule selects the set of envy free allocations whenever it is non empty.*

Proof. $\forall \boldsymbol{p} \in \mathbb{R}^{N_h}$, $\forall i \in N_h$, $\sum_j \epsilon_{ij} \geq 0 \iff i$ does not envy any other agent. Therefore, $\sum_{i,j} \epsilon_{ij} = 0 \iff \boldsymbol{p}$ is envy-free. □

Having described the concept of PEXIC, the algorithm for signal phasing optimization using individual vehicle information, and the subsequent compensatory pricing mechanism in detail, we now proceed to modeling the scheme on a simulated single intersection context.

4.5 Simulation Results

For our simulation studies, we use an intersection with eight phases that is modelled using the microscopic simulation software, PARAMICS (Quadstone Ltd, 2010). Every left turn movement has one lane, and the through movements have two lanes each. Left turn lanes have a bay that is 250 ft. long. The approach speed is 40 mph and vehicles are released at that speed to simplify control delay measurements. Each approach is 2000 ft. long.

Each simulation experiment is run for three hours. Experiments differ in the critical volume of the intersection, the seed used and the percent of vehicles currently running late (i.e., expected to be late at the destination). We run six seeds and three values for percentage of late vehicles (PLV): 0%, 7.5% and 15%. Each experiment is run for both delay minimization (anonymous control, FCFS) and our control (efficient cost minimization with minimum envy payments). The volumes used for the simulation cases are shown in Table 4.1. Arrival flows differ in direction and movement as follows. Northbound and southbound movements are 75% of Eastbound and Westbound movements. Left turn flows are two thirds of their adjacent through movements. This second relation comes from the maximum recommended values of flow per lane in the Highway Capacity Manual (National Research Council Transportation Research Board, 2010). The highest flow row corresponds to the maximum recommended limits of flows per lane in HCM, 450 veh/h/lane in through movements and 300 veh/h/lane in left turns. These relations give the following flows in veh/h:

Critical Volume	EW Through	NS Through	EW Left	NS Right
150	50	38	33	25
450	150	113	100	75
750	250	188	166	125
1050	350	263	233	175
1350	450	338	300	225

Table 4.1: Critical volumes per approach

Vehicles are randomly assigned a lane on their entry and are not allowed to change lane, in order to simplify arrival prediction. Arrivals are predicted on projected trajectories at the approach speed. If the simulated speed is below the queuing speed threshold (below which the vehicle is assumed to be in a queue), the vehicle is considered in the vertical queue. Left turn vehicles are also randomly assigned an approach arrival lane, and become aware of the intersection by a sign post at 200 ft. before the approach widens with the left turn bays. All parameters used are PARAMICS default values. Vehicles' control delay is measured 600 ft. downstream of the intersection which is an observed distance far enough for vehicles to again reach their desired speed (the speed at which they entered the system).

4.5.1 Delay and cost results

Table 4.2 displays the final average cost per vehicle under the proposed control (shown as PEXIC) and under delay minimization (shown as "Min-Delay"). The table shows the results for the three PLVs (percentage of late vehicles). Generally, our control reduces travel cost up to 20% and it decreases as the critical flow increases or the percentage of late vehicles decrease. We observe that at higher flows the efficiency increase (i.e., decrease in cost) is lower than in lower flows, but still positive. This suggests that the policy does scale very well on arrival flow. The weaker efficiency gains at higher flows are due to the number of vehicles per approach increasing, thereby bringing the average vehicle VDS per approach closer to the population mean. This means that extending one phase or a barrier group of the dual ring control does not lead to significant cost reductions, since prioritizing any of the approaches would bring similar benefits. The only exception is the lowest critical flow value which has a lower efficiency increase. Reducing the critical volume up to very low values decreases the probability of intersection conflict, leading to zero delay for most vehicles regardless of the control.

	av. cost PEXIC ($/veh.)			av. cost Min-Delay ($/veh.)			av. cost incr. (%)		
Volume (veh./h)	15% PLV	7.5% PLV	0% PLV	15% PLV	7.5% PLV	0% PLV	15% PLV	7.5% PLV	0% PLV
150	0.012	0.01	0.009	0.015	0.012	0.009	-18.4	-14.03	-2.63
450	0.027	0.023	0.019	0.034	0.027	0.02	-20.21	-15.33	-5.33
750	0.043	0.036	0.029	0.05	0.041	0.03	-14.28	-12.58	-4.21
1050	0.063	0.053	0.042	0.071	0.057	0.043	-10.69	-7.7	-1.87
1350	0.098	0.08	0.061	0.103	0.082	0.063	-4.74	-3.25	-2.41

Table 4.2: Average costs and cost savings as a function of critical volume

As expected, the efficiency increase is largely influenced by the skewness of the VDS distribution, which is more for higher percentages of late vehicles. This is clear in the 0% PLV columns of Table 2. They show that the VOT distribution, which has the shortest tail, leads to lower efficiency gains closer to zero throughout the volume range. We conclude that symmetry in the distribution evens up any long run potential efficiency gains from modifying the phasing from the minimum delay phasing.

Table 4.3 shows how minimizing an objective function other than delay naturally increases delay. However, delays from cost minimization show an interesting pattern: the percentage increase remains approximately constant regardless of the critical volume. This pattern is present in both levels of percentage of late vehicles (PLV) meaning that the negative consequences of the policy are bounded. Precaution needs to be taken for higher flows or in case the intersections are closely located. Additional delays could produce spillbacks and gridlock. In case of sustained volumes close to saturation condition, the policy may have to be deactivated and switched to delay minimization, but such eventualities require further study.

Volume (veh/h)	Av. delay Min-Delay (s)	Av. delay PEXIC (s)			Av. delay increase (%)		
		PLV 15%	PLV 7.5%	PLV 0%	15% PLV	7.5% PLV	0% PLV
150	3.61	4.26	4.12	3.96	17.85	13.98	9.67
450	8.24	9.36	9.12	8.66	13.59	10.61	5.05
750	12.25	14.09	13.6	12.77	15.04	11.11	4.29
1050	17.31	19.96	19.4	18.15	15.34	12.1	4.87
1350	25.23	29.19	28.51	25.85	15.74	13.03	2.55

Table 4.3: Delays and delay increase with respect to critical volume

Fig.4.3 shows the results from one of the experiments, for critical volume equal to 750 veh/h and a PLV of 15%. It exemplifies the discussion made above. The upper left figure shows the delay savings over time for every single vehicle colored by value of time intervals. We can observe that most of the higher VDS vehicles have positive delay savings, while most of the lower VDS vehicles lie in the negative part. By multiplying the former individual delays by their respective individual VDS we obtain the individual cost savings (upper right figure). We observe how the cost differences are centered on zero, having low VDS vehicles a lower variance and the high VDS vehicles a higher one, but positively skewed as the VDS increases.

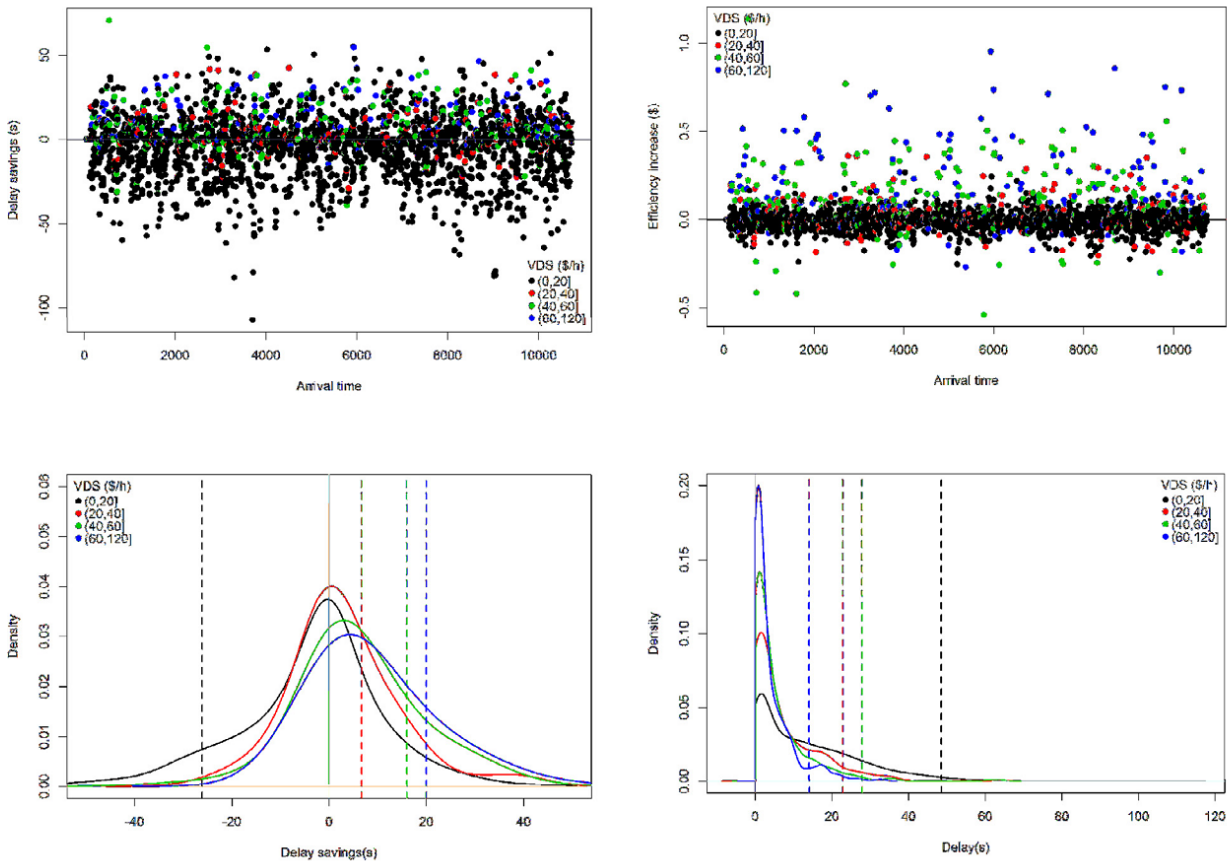


Figure 4.3: Upper left figure: Individual delay savings w.r.t. Min-Delay. Upper right figure: individual cost savings w.r.t. Min-Delay. Lower left figure: density of delay savings w.r.t to Min-Delay. Lower right figure: delay pdfs. All plots represent a single seed of problem instance flow = 750 veh/h and 15% PLV.

The lower right graph of Fig.4.3 shows how the cost minimization rule successfully assigns lower delays to high valuation vehicles and higher delays to the rest. The vertical dashed lines show the average delay savings and delays for every valuation range. This phenomenon is what leads to the final average efficiency increase. This can be better appreciated in the lower left graph, which shows the delay savings distribution as a function of the value of delay savings. As the flow increases, a greater percent of high valuation vehicles experience negative cost savings, neutralizing

Table 4.4 shows the efficiency and delay increases disaggregated by valuation intervals, for the three PLV values. We observe that the vehicles which are not late, i.e. the ones with a valuation lower than \$20/h, are the only valuation group that sees their costs increase, compared to when delays are minimized. Percentage increase in efficiency (decreases in costs) have similar magnitudes as decreases in delay. Furthermore, we see how for all the valuation categories, the increase in delay is similar to the increase in cost except for the lowest valuation category. Finally, minimizing costs does not lead to a significant efficiency increase over delay minimization. Evidence suggests that a minimum skewness degree in valuation distribution is needed. For this reason, we do not analyze further the 0% late vehicles case.

Considering now the critical flow dimension, the percent cost increase for the lower valuation vehicles remains constant close to +20%, while higher valuation vehicles experience a larger efficiency increase as their valuation increases. Increasing the critical volume, though, reduces the efficiency increase for these valuations, as we observed in previous discussions.

		Delay PEXIC (s)			Delay inc. (%)			Cost inc. (%)		
Vol. (veh/h)	VDS ($/h)	PLV 15%	PLV 7.5%	PLV 0%	PLV 15%	PLV 7.5%	PLV 0%	PLV 15%	PLV 7.5%	PLV 0%
150	(0,20]	4.7	4.3	4.0	31.0	19.9	9.7	20.20	7.7	-2.6
	(20,40]	2.0	1.6	-	-44.9	-56.0	-	-49.4	-61.6	-
	(40,60]	1.7	1.4	-	-54.1	-60.4	-	-58.7	-66.6	-
	(60,150]	1.1	1.0	-	-68.7	-72.6	-	-62.4	-62.7	-
450	(0,20]	10.3	9.6	8.7	25.2	16.3	5.1	15.8	5.3	-5.4
	(20,40]	5.1	4.5	-	-37.8	-46.0	-	-38.6	-48.3	-
	(40,60]	3.5	2.8	-	-57.1	-66.0	-	-59.2	-66.8	-
	(60,150]	2.8	2.1	-	-65.7	-74.4	-	-67.7	-73.7	-
750	(0,20]	15.4	14.3	12.8	25.3	16.4	4.3	18.1	8.9	-4.2
	(20,40]	8.3	7.2	-	-32.4	-41.2	-	-33.9	-44.1	-
	(40,60]	6.9	5.1	-	-44.1	-58.8	-	-45.3	-59.4	-
	(60,150]	4.6	3.7	-	-62.9	-70.2	-	-64.2	-72.7	-
1050	(0,20]	21.6	20.2	18.2	24.9	16.8	4.9	20.7	11.6	-1.9
	(20,40]	12.8	11.3	-	-25.8	-34.5	-	-28.7	-37.2	-
	(40,60]	10.1	8.3	-	-41.5	-51.8	-	-42.0	-51.4	-
	(60,150]	7.6	7.00	-	-56.0	-59.6	-	-57.4	-62.7	-
1350	(0,20]	31.1	29.5	25.9	23.2	17.0	2.5	19.6	13.0	-2.5
	(20,40]	21.2	19.8	-	-15.8	-21.5	-	-16.9	-22.5	-
	(40,60]	17.7	14.2	-	-29.7	-43.6	-	-30.1	-43.2	-
	(60,150]	14.3	11.1	-	-43.5	-55.9	-	-44.0	-56.5	-

Table 4.4: Delays, delay increases, cost, cost increase as a function of volume, VDS and PLV.

It is important to also look into inequities across the vehicles and make sure that no extreme case of benefit-starvation occurs. Thus, Fig. 4.4 and Fig. 4.5 show the delays per vehicle for each of the 8 phases (movements). Fig. 4.4 shows the delays in the case of optimizing the phases for minimum total delay, to which each of the cases of cost minimization shown in Fig. 4.5 (for PLVs of 7.5 and 15%) can be compared. The cost minimization does not cause undue benefits or costs to any phase in comparison to delay minimization, as the patterns are extremely similar when the plots in Fig. 4.5 are compared to Fig. 4.4. Generally, the left turn phases and minor (lower flow) phases have higher average costs and delays than the through phases and major phases. As we increase the flow, left turn movements see delay increases that are substantially more than the through movements. Altering the percentage of late vehicles does not have any influence on the delay difference across phases, rather it just slightly increases all of them homothetically, as a comparison of the plots in Fig. 4.5 shows.

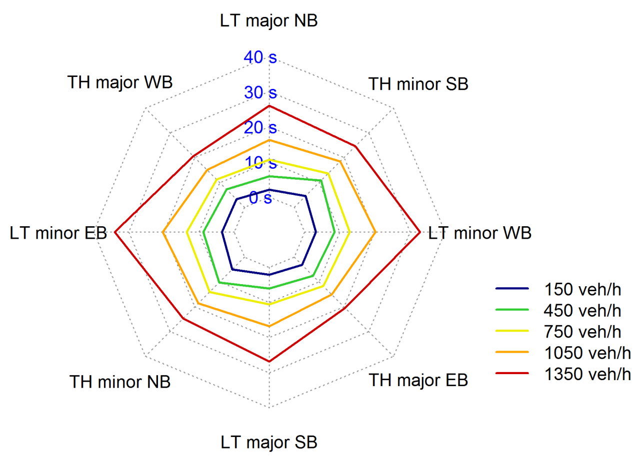


Figure 4.4: Av. delays per phase with respect to critical flow under delay minimization

4.5.2 Pricing results and utilities: welfare analysis.

The following results are based on the logistic weights (4.23) as described earlier, with parameters a equal to 12 and b equal to 6, which were the minimum values that were

Figure 4.5: Average PEXIC delays per phase with respect to critical flow for a PLV of 7.5% (left) and 15% (right).

found to provide Pareto efficiency in preliminary studies. In real-world applications such weights can be calibrated or learned. These were values that proved the control policy to compensate enough to be Pareto efficient with respect to delay minimization, but not redistributive enough to allow arbitrage, that is, giving low value vehicles an incentive to drive to collect payments from other vehicles, an eventuality that must be avoided.

Fig. 4.6 shows the average price with respect to the experienced delay. The resulting fitted curves from here and beyond are estimated by locally weighted regressions (Cleveland et al., 1992) with default parameters from the loess package in R (R Development Core Team, 2014). As expected, prices decrease with the delay. Prices are negative for higher delays due to the budget balance condition. For a particular delay value, the average price increases with the critical volume. Higher critical volumes show higher average delays, thus translating the curves to the right. Since the sum of prices is balanced, increasing the percent of vehicles rotates the pricing curves clockwise. The VDS distribution is now more skewed to the right and high VDS vehicles experience a higher delay than before since it becomes more costly to reduce their delay, as we have seen in the previous section.

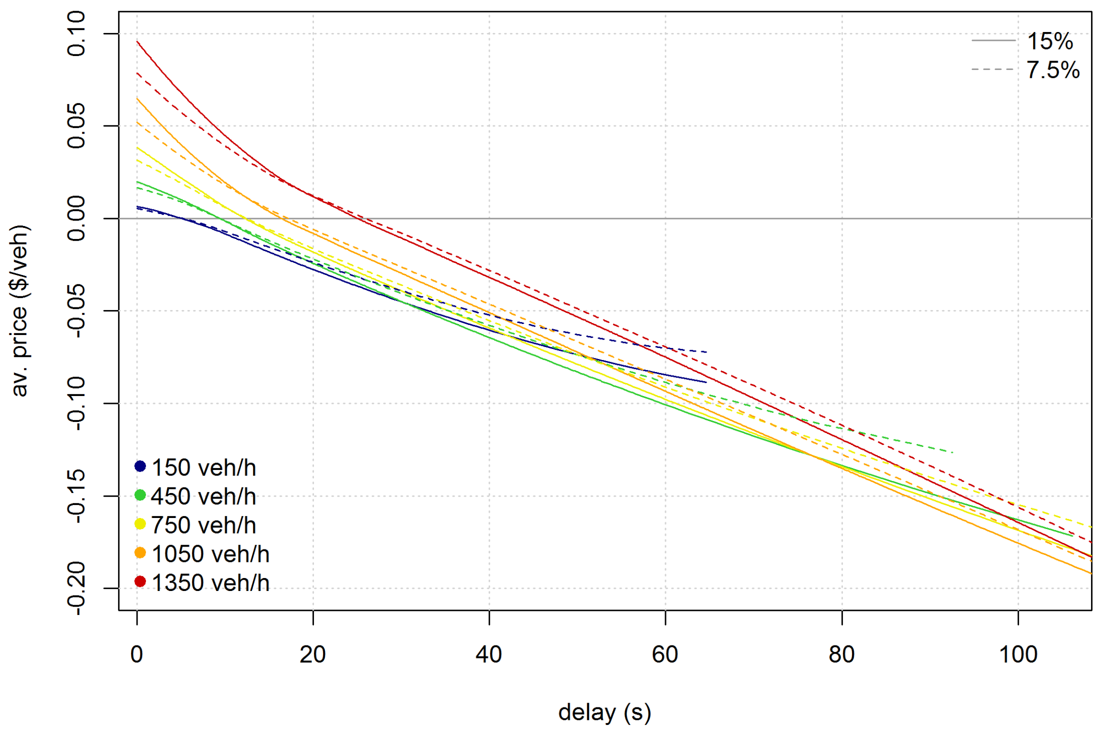


Figure 4.6: Average price vs delay for different volume levels and percentages of late vehicles (PLV)

Fig. 4.7 shows how the average price paid per vehicle against their VDS. Prices naturally increase with the VDS, and are negative for the lowest VDS and positive for the higher ones. The cut-off points are all close to the population VDS mean, due to budget balancing. As critical volume increases, prices becomes more extreme, which leads to a more convex price curve. In this sense, the weighted additive criterion successfully manages to charge higher average prices to higher VDS vehicles and even higher prices at higher flows, since the efficiency increase is more expensive at high volumes. Finally, the pricing is sensitive to the percentage of late vehicles: high VDS vehicles pay more since now they need to compensate for low valuation vehicles' higher delays.

Fig. 4.8 compares the average utility increase over using an FCFS rule (i.e. delay minimization is used as the benchmark) as a function of the VDS, θ. We can conclude that

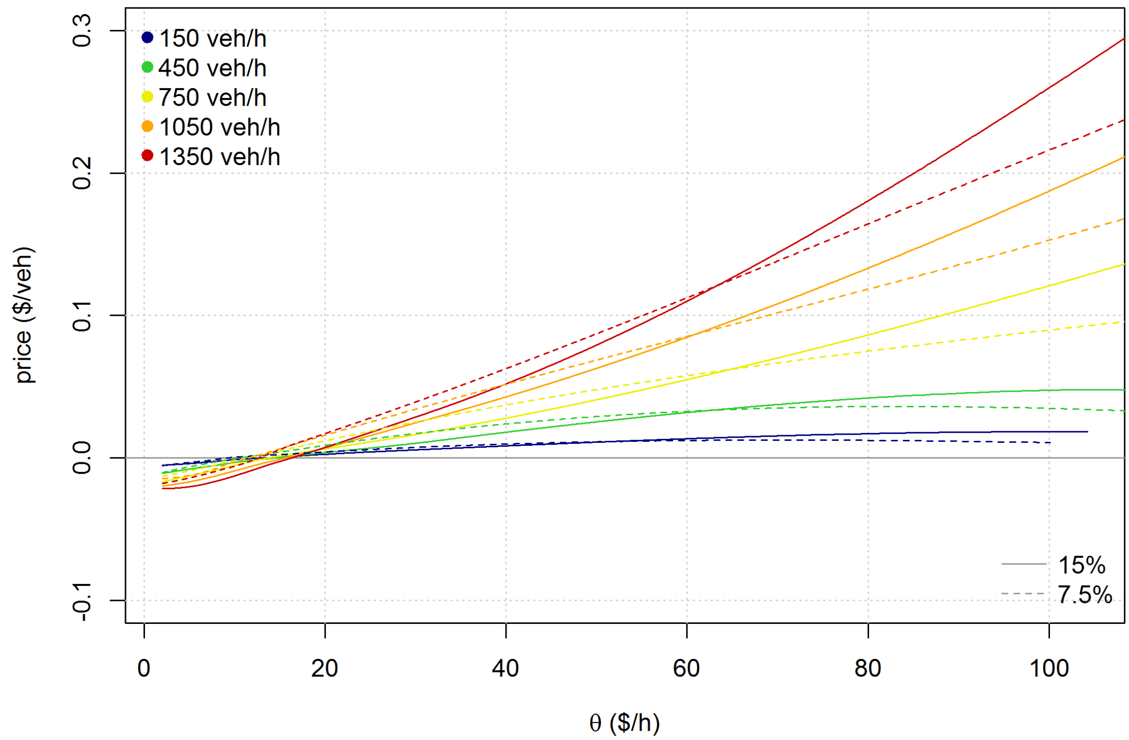


Figure 4.7: Average price versus value of delay savings (VDS), for different volume levels and percentages of late vehicles (PLV)

the policy is Pareto-improving, compared to delay minimization. That is, all vehicles, regardless of their VDS and average incoming flow, are better off, on average, under the new control system than under delay minimization. For the lower types, the non-late vehicles, the average utility increase is weakly positive, meaning that minimum-envy pricing is capable of compensating for the increase in delay in comparison to traditional delay minimization. For higher types, the utility difference gets larger, which means that the marginal utility under our policy is larger than the average delay under delay minimization, which is a constant regardless of the vehicle type.

Critical volume generally has a positive effect on the average utility increase for higher VDS, since the marginal price w.r.t. critical volume is smaller than marginal cost difference between PEXIC and delay minimization. This can be seen from the slopes of

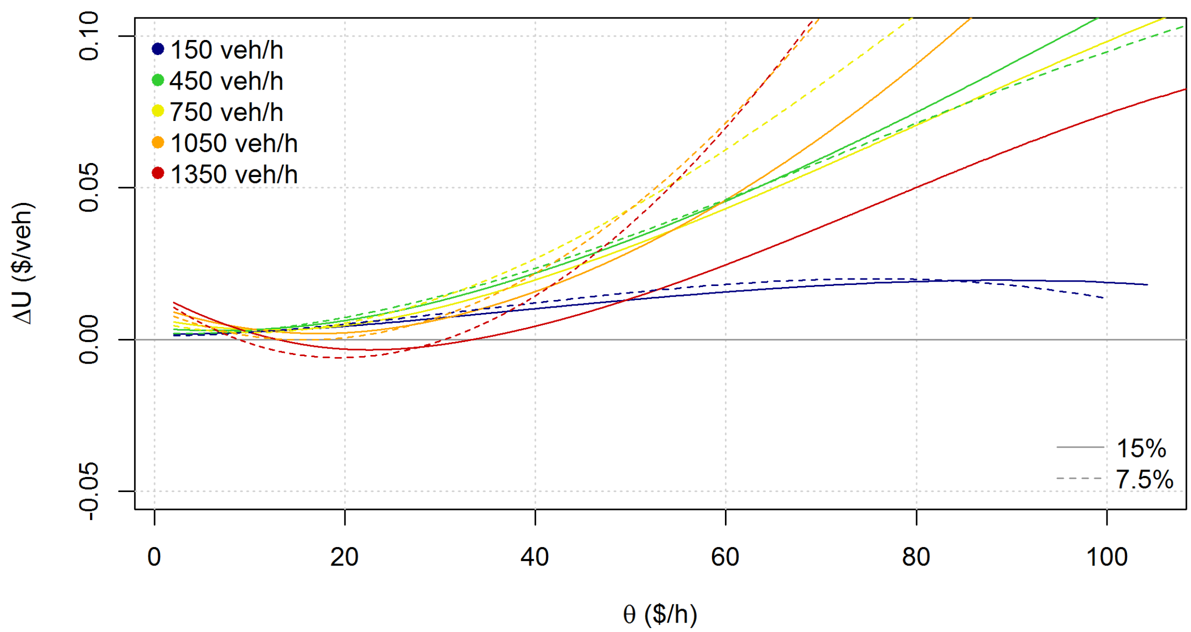


Figure 4.8: Average utility increase versus value of delay savings (VDS), for different volume levels and percentages of late vehicles (PLV)

the curves, as the solid curves always have higher slope than the corresponding dashed curves. The percentage of late vehicles (PLV) has less effect on the utility increase for low critical volumes. The prices are insensitive to this variable and the delays stay close to zero. For larger volumes, it is more costly now to compensate the bulk of low VDS vehicles since now the experience a higher delay.

Vehicles with the highest VDS experience higher increases in utility. This indicates that there is room for more redistributive pricing policies. However, the VDS distribution is highly skewed, in that a few high value vehicles need to compensate for a large number of non-late low-VDS vehicles. To better assess this limitation, Table 4.5 shows the average utility increase for all, non-late ($\theta < \$20$/h) and late ($\theta \geq \$20$/h) vehicles. For instance, for every extra dollar that is subtracted from every high VDS vehicle, low value vehicles' utilities increase just by \$0.17 (for x=15%).

PLV (%)	Vol. (veh/h)	av. U increase ($/100 veh)		
		all values	low value	high value
15 %	150	0.21	0.02	1.21
	450	0.60	0.15	3.15
	750	0.63	0.21	3.05
	1050	0.69	0.28	3.02
	1350	0.44	0.25	1.59
7.5%	150	0.13	0.00	1.75
	450	0.35	0.09	3.70
	750	0.45	0.09	4.98
	1050	0.39	0.04	4.80
	1350	0.24	-0.09	4.42

Table 4.5: Summary of average utility increases

A more redistributive policy could lead to negative utility increases for the late vehicles while further utility increase of non-late vehicles would not be that substantial. Nevertheless, higher value of delay savings maybe typically tied to vehicles which proceed to activities of high added value. In fact, offering substantial incentives to this type of participants may be social beneficial anyway. We could even argue that this policy therefore provides further economic benefits beyond travel cost reduction.

Finally, a higher degree of compensation could induce low VDS users to drive to collect payments and increase congestion and VMT. We analyze this situation by observing the average agents' utilities on all the scenarios. Users will not be able to collect payments if their utilities are non-positive. Table 4.6 shows that for all the VDS ranges, utilities remain negative for all the variable range. This essentially means that the disutility from such additional driving and waiting is not offset by the received payment. We conclude that the policy is not beneficial for anyone to collect payments for personal profit.

PLV (%)	Vol. (veh/h)	θ ($/h)					
		(0,5]	(5,10]	(10,20]	(20,40]	(40,70]	(70,150]
15 %	150	-0.003	-0.007	-0.014	-0.023	-0.034	-0.047
	450	-0.006	-0.016	-0.030	-0.057	-0.081	-0.101
	750	-0.006	-0.023	-0.047	-0.091	-0.140	-0.190
	1050	-0.009	-0.032	-0.068	-0.141	-0.214	-0.306
	1350	-0.016	-0.048	-0.101	-0.221	-0.340	-0.522
7.5%	150	-0.003	-0.008	-0.013	-0.019	-0.030	-0.029
	450	-0.007	-0.017	-0.030	-0.053	-0.068	-0.081
	750	-0.009	-0.025	-0.046	-0.088	-0.120	-0.151
	1050	-0.013	-0.035	-0.068	-0.134	-0.195	-0.268
	1350	-0.003	-0.007	-0.014	-0.023	-0.034	-0.047

Table 4.6: Vehicles' average utility as a function of VDS, PLV and volume

4.5.3 Fairness results

In this section, we analyze how the proposed control policy is fairer than delay minimization (Min-delay). We measure fairness by using two different metrics composed of the resulting individual envy values. These metrics are calculated for each of the two policies. The two measures differ on how the envy terms are aggregated after every rolling horizon optimization h. The first one, ϵ^{max}, calculates the average of the maximum individual rolling horizon envy terms ϵ_{ijh}, while the second, ϵ^{sum} calculates the average of the sum of the rolling horizon envy terms:

$$\epsilon^{max} = \frac{1}{N}\sum_{i}\max_{I_h \ni i}\max_{j}\epsilon_{ijh} \tag{4.28}$$

$$\epsilon^{sum} = \frac{1}{N}\sum_{i}\sum_{I_h \ni i}\max_{j}\epsilon_{ijh} \tag{4.29}$$

In Table 4.7 we observe how our control policy is generally fairer than Min-Delay for all critical volumes and percent of late vehicles, over most of the vehicle VDS range. Only in the case of saturation critical flow, does the policy not manage to compensate low VDS vehicles enough. Therefore, our policy successfully increases satisfaction for late vehicles by providing them shorter delays, while at the same time compensates lower VDS vehicles for their longer delays. We observe that regardless of the metrics and control policy, envy increases with flow and percentage of late vehicles. To explain these results, we rearrange (4.25) for both policies to obtain individual envy intensity expressions:

$$\epsilon_{ih}^{Min-delay} = \theta_i \max(d_{ih} - \min_{I_h \ni j \neq i} d_{jh}\ ,\ 0) \tag{4.30}$$

$$\epsilon_{ih}^{PEXIC} = \max\Big(\theta_i(d_{ih} - d_{jh}\Big) + p_{ih} - p_{jh} + \pi_{ih} - \pi_{jh},\ 0) \tag{4.31}$$

Generally, if the percentage of late vehicles increases, more vehicles have a higher θ_i and the resulting final objective value, the total average envy, increases. In the Min-Delay case, delays are independent of VDS. Late vehicles will naturally experience more envy due to larger θ_i values. Increasing the critical volume increases vehicles' delay, causing a larger difference between vehicle i's delay and the minimum delay. This distances the delay assignment from a swap monotone delay assignment, which would be envy-free with zero envy intensity is zero.

In comparison, (4.31) now incorporates differences in prices, which allows lower levels of envy. Interpretation of the sensitivity analysis is less intuitive, since the prices are constrained by budget balancing and max operators. In general terms, late vehicles are charged higher prices and their envy terms need to increase to satisfy the envy conditions, especially when they are envying low-VDS vehicles which are positively compensated. The opposite situation happens then for those low-VDS vehicles: they will not envy the rest of vehicles and the ϵ_i terms will get close be zero.

The policy is fairer for everybody, but especially for high VDS vehicles. Only in the case of critical flow close to saturation ratio does the policy not outperform delay minimization on fairness. This, added to the fact that efficiency increase at this volume level is close to zero suggests that the policy should be switched to delay minimization when the intersection reaches saturation.

Metrics	PLV	Volume (veh/h)	Envy Min-Delay ($/veh)			Envy PEXIC ($/veh)			Envy decrease (%)		
			all	Low	high	all	low	high	all	low	high
Max	15%	150	0.07	0.03	0.06	0.03	0.02	0.02	-50.4	-35.0	-65.6
		450	0.19	0.10	0.16	0.13	0.09	0.08	-30.2	-9.2	-52.0
		750	0.31	0.16	0.25	0.24	0.15	0.15	-23.2	-5.9	-41.6
		1050	0.48	0.25	0.38	0.40	0.26	0.24	-15.5	4.3	-36.4
		1350	0.69	0.36	0.55	0.67	0.43	0.41	-3.4	17.7	-26.6
	7.5%	150	0.06	0.04	0.05	0.02	0.02	0.01	-54.8	-45.6	-75.2
		450	0.15	0.11	0.13	0.10	0.08	0.05	-35.2	-25.1	-59.9
		750	0.25	0.18	0.22	0.18	0.14	0.10	-29.4	-19.4	-52.0
		1050	0.38	0.27	0.32	0.29	0.23	0.18	-23.0	-13.6	-44.4
		1350	0.56	0.39	0.47	0.49	0.38	0.29	-12.7	-2.3	-37.5
Sum	15%	150	0.13	0.06	0.11	0.05	0.03	0.03	-63.4	-52.9	-74.0
		450	0.46	0.23	0.38	0.25	0.17	0.14	-44.8	-28.8	-61.5
		750	0.80	0.41	0.65	0.51	0.32	0.31	-36.4	-21.5	-52.4
		1050	1.16	0.59	0.94	0.85	0.54	0.51	-26.8	-8.4	-46.0
		1350	1.56	0.82	1.25	1.35	0.87	0.81	-13.5	6.2	-35.1
	7.5%	150	0.10	0.07	0.10	0.03	0.03	0.02	-66.5	-59.7	-81.6
		450	0.36	0.25	0.31	0.18	0.15	0.10	-49.4	-41.3	-68.8
		750	0.65	0.45	0.56	0.38	0.30	0.21	-41.8	-32.7	-62.3
		1050	0.92	0.64	0.78	0.61	0.48	0.36	-33.6	-24.8	-53.7
		1350	1.25	0.89	1.05	0.97	0.78	0.56	-22.5	-12.6	-46.5

Table 4.7: Average envy per vehicle as a function of PLV, volume and VDS segment (low, high or all).

4.6 Extension: PEXIC under CESEI envy-minimizing criteria family

We will now explore the behavior of the CESEI criterion on the PEXIC control system. CESEI criterion will be evaluated for the values $\rho = \{1, 2, \infty\}$. The β_j terms are set to 1 and the γ_i will follow the same specification as in the previous section.

$$\gamma_i = 1/(1 + a\exp(-b(\overline{\theta}_h - \theta_i)/(\overline{\theta}_h - \underline{\theta}_h))) \; \forall i \in I_h \tag{4.32}$$

PEXIC with CESEI criteria results are compared against standard delay minimization.The sensitivity analysis is done on the critical volume of the intersection and the ρ parameter. Originally, Lloret-Batlle and Jayakrishnan (2016) used the CESEI criterion with $\rho = 1$ and $\beta_j = 1 \; \forall j \in I$. The substitution factor was set to one, to model perfect substitution and allow a fast linear programming formulation. β_j are set to one since there is no interest on weighting the envy relations by the envied agent j but by the envious agent i. Instead, the weights $\gamma_i \neq 1$ are set to follow a logistic specification, leading to monotonic decreasing weights with respect to VDS. This functional form makes the lower VDS-associated envy terms more costly than the high ones. Consequently, envy terms of the low value vehicles are minimized further than those of the high value vehicles, leading to higher average payments for the latter. The logistic specification was chosen for its flexibility and simplicity, and also for its suitable bounds.

The resulting optimization program, ρ-CESEI_DYN for any rolling horizon h is:

$$\min \sum_{i \in I_h} \gamma_i \left(\sum_{j \neq i} \left(\epsilon_{ij}(\mathbf{p}) \right)^{\rho} \right)^{1/\rho} \tag{4.33}$$

s.t

$$\epsilon_{ij} - \theta_i d_i - p_i - \pi_i^{h-1} \geq -\theta_i d_j - p_j - \pi_j^{h-1} \; \forall i, j \in I_h, i \neq j \tag{4.34}$$

$$\sum_i p_i = 0 \tag{4.35}$$

$$\epsilon_{ij} \geq 0 \; \forall i, j \in I_h, i \neq j \tag{4.36}$$

Here, (4.33) is the CESEI objective function, (4.34) are the dynamic relaxed envy-free conditions, (4.35) is the budget balance constraint and (4.36) are the non-negativity envy term constraints.

Since the presented model results in joint actions among users in consuming the supply (i.e., signal time), we now examine the fundamental interactions in this collaborative consumption scheme using the framework in Chapter 1. For PEXIC, the operator objective O is efficiency maximization and envy minimization with CESEI criterion. The regulation environment E is minimum envy-freeness, that follows the CESEI criterion and assumes honest preference elicitation. Since the policy is budget balanced, operator's resources R is the empty set. The allocation and price rule $[T, F]$ is Algorithm 1 and the collaborative exchange procedure is the execution of Algorithm 1.

We next introduce two measures of envy in dynamic problems: (average) maximal dynamic envy and (average) accumulated dynamic envy. The first one calculates the average of the maximum individual rolling horizon envy terms ϵ_{ijh}, while the second calculates the average of the sum of the rolling horizon envy terms.

Definition 4.2. *(average) Maximal dynamic envy measure:*

$$\epsilon^{max} = \frac{1}{N} \sum_i \max_{h:\, I_h \ni i} \{\max_j \{\epsilon_{ijh}\}\} \tag{4.37}$$

Definition 4.3. *(average) Cumulative dynamic envy measure:*

$$\epsilon^{sum} = \frac{1}{N} \sum_{i} \sum_{h:\, I_h \ni i} \{\max_{j}\{\epsilon_{ijh}\}\} \tag{4.38}$$

These measures are in line with the interpretation that the actual envy intensity an agent i experiences is the maximum of all envy intensities resulting from all the envy comparisons with the rest of agents for all the rolling horizons. That is, the envy intensity is the maximum ϵ_{ijh} that an agent experiences during his presence inside the system. The second measure differs from the former in which the envy intensity perceived is the sum of all the maximum envy intensities the agent experiences during the his presence into the system.

We proceed next to compare the resulting levels of envy of the PEXIC system under CESEI criterion (4.33) against standard delay minimization. First, though, we reorganize the envy-free equations (4.34) for easier interpretation of results:

$$\epsilon_{ij}^{PEXIC} = \max\left\{\theta_i(d_i - d_j) + p_i - p_j + \pi_i^{h-1} - \pi_j^{h-1},\ 0\right\} \tag{4.39}$$

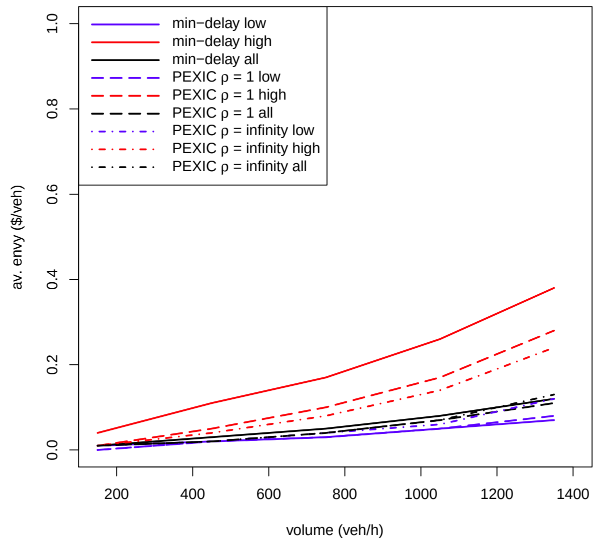
In the delay minimization case, envy intensity values are obtained by solving the same problem (4.33-4.36) but constraining the payments $p_i, \pi_i^{h-1}\ \forall i \in I_h\ \forall h$ to zero:

$$\epsilon_{ij}^{mindelay} = \theta_i \max\left\{d_i - \min\left\{d_j\right\},\ 0\right\} \tag{4.40}$$

Equation (4.39) incorporates differences in prices, which allows lower levels of envy. Generally, high-VDS vehicles are charged higher prices and their envy terms need to increase to satisfy the envy conditions, especially when they are envying low-VDS vehicles which are positively compensated. The opposite situation happens then for those low-VDS vehicles: they will not envy the rest of vehicles and the ϵ_{ij} terms will get close to zero.

Figure 4.9 shows the results for the first two envy measures: maximal envy intensity (on the left) and cumulative envy intensity (on the right). Results are shown for three cases: the delay minimization scenario ("min-delay"), PEXIC with $\rho = 1$ and PEXIC with $\rho = \infty$. Results with $\rho = 2$ consistently lay very close to $\rho = 1$ and are not shown in the figures, for clarity. The simulation configuration is the same as in Lloret-Batlle and Jayakrishnan (2016). Envy levels are displayed in three groups: vehicles with VDS lower that \$20/h ("low"), vehicles with VDS higher than \$20/h ("high") and all vehicles together ("all"). In the delay minimization ("min-delay") scenario, delays are independent of VDS θ_i.

We observe how high VDS vehicles naturally experience more envy due to larger θ_i values. Increasing the critical volume naturally increases vehicles' delay, makes it more difficult to find a phasing which resembles an ideal (envy-free) monotone delay assignment with no envy, leading to increased envy levels. When $\rho = \infty$, which corresponds to a minimax criterion, PEXIC does not manage to compensate low value vehicles enough and their envy levels, especially when measured in the cumulative metrics (right figure) are higher than in delay minimization. However, with $\rho = 1$, CESEI criterion does manage to be fairer than delay minimization regardless of the VDS value, justifying the usefulness of this criteria for dynamic envy-freeness.



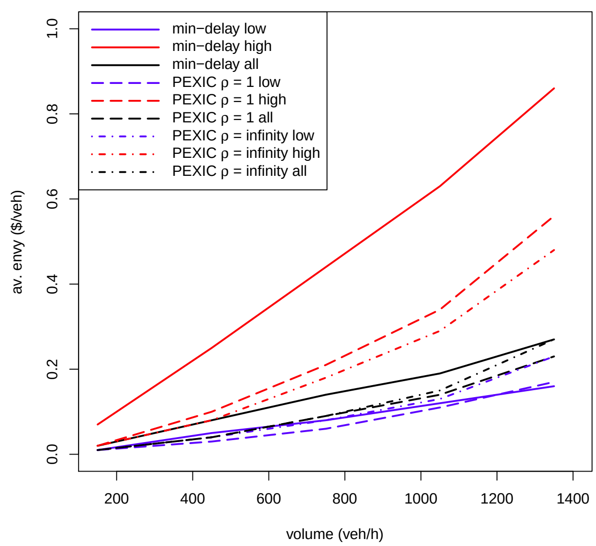


Figure 4.9: Maximal and cumulative average envy per vehicle as a function of critical volume, CESEI ρ parameter and VDS segment (”low”: < \$20/h, ”high”: > \$20/h or ”all” vehicles) ”min-delay” stands for delay miminzation control.

Figure 4.10 shows the average utility increase of PEXIC control system (Algorithm 1) with regard to delay minimization with regard to vehicles’ VDS, for different critical volume levels and $\rho = \{1, 2, \infty\}$. We can observe that for a minimax criteria, $\rho = \infty$, not enough utility is being transferred from high VDS vehicles to low VDS vehicles. However, for low values of ρ, the criterion manages to transfer enough utility, and making PEXIC a Pareto improving (and optimal) with regard to delay minimization, for any volume regime and vehicle VDS segment.

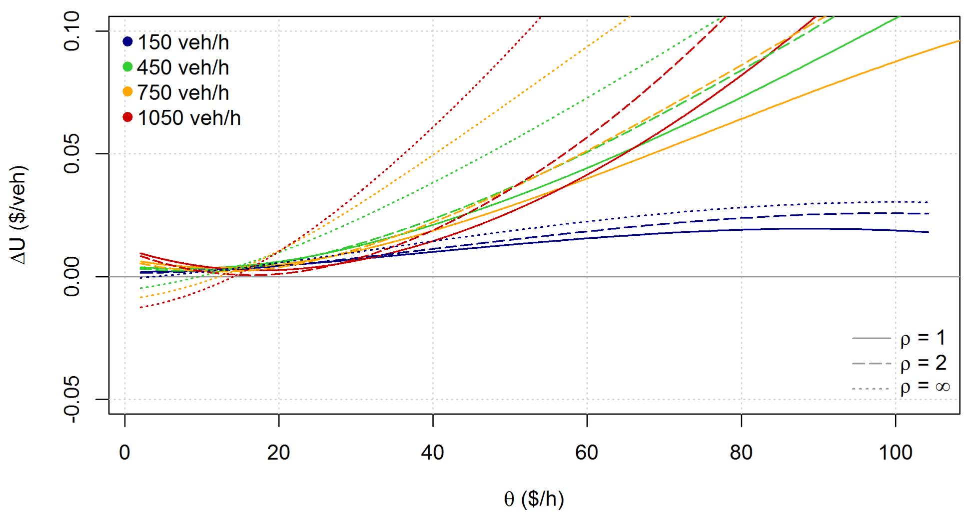


Figure 4.10: Maximal and cumulative average envy per vehicle as a function of critical volume, CESEI ρ parameter and VDS segment (”low”: < \$20/h, ”high”: > \$20/h or ”all” vehicles) ”min-delay” stands for delay miminzation control.

4.7 PEXIC and Incetive Compatibility

This section further expands the contents of the original PEXIC article (Lloret-Batlle and Jayakrishnan, 2016). Here, we empirically explore the incentive-compatible character of the minimum-envy prices found in the former section.

To compare the minimum envy prices with Bayesian Incentive-Compatible (BIC) prices, we will make use of Myerson's characterization theorem for linear environments (Mas-Colell et al., 1995), described in the previous chapter. We again define such an environment.

Definition 4.4. *Linear environment:*

1. *Each type θ_i lies in $\Theta_i = \left[\underline{\theta}_i, \overline{\theta}_i\right]$ $\forall i \in \{1, \ldots, n\}$, where $\underline{\theta}_i < \overline{\theta}_i$.*
2. *Agents' types are statistically independent $\phi(\bullet) = \times_i \phi_i(\bullet)$.*
3. $\phi_i(\theta_i) > 0 \quad \forall \theta_i \in \left[\underline{\theta}_i, \overline{\theta}_i\right]$ $\forall i \in \{1, \ldots, n\}$
4. *Each agent i's utility function has the form: $u_i(x, \theta_i) = \theta_i v_i(k) + t_i$*

It is easy to check that PEXIC environment is indeed linear.

Theorem 4.2. *(Myerson's Characterization Theorem)*
In a linear environment, a social choice function $f() = (k(), t_1(), \ldots, t_n())$ is BIC if and only if, $\forall i \in \{1, \ldots, n\}$:

1. $\overline{v}_i()$ is non-decreasing.

$$2. U_i(\theta_i) = U_i\left(\underline{\theta}_i\right) + \int_{\underline{\theta}_i}^{\theta_i} \overline{v}_i(s)\, ds \quad \forall \theta_i \in \Theta_i$$

Where $\overline{v}_i(s)$ is the average valuation function for agent i when he reveals type s. $U_i(\theta_i)$ is the expected utility of user i when he has type θ_i and all agents reveal their types truthfully, including i.

By replacing the utility function specification into the second condition keeping the expectation, we can isolate the average transfer $\overline{t}_i(\theta_i)$.

$$\overline{t}_i(\theta_i) = \overline{t}_i\left(\underline{\theta_i}\right) + \underline{\theta_i}\overline{v_i}\left(\underline{\theta_i}\right) - \theta_i\overline{v_i}(\theta_i) + \int_{\underline{\theta_i}}^{\theta_i} \overline{v_i}(\theta_i)\,ds \qquad [EP]$$

$\overline{t}_i(\theta_i)$ describes a family of price functions, with a free constant $\overline{t}_i\left(\underline{\theta_i}\right)$.

To (asymptotically) compare both PEXIC prices and BIC prices, we are going to follow this procedure, which respects the microeconomic interpretation of $[EP]$:

1. Split VDS values into the following intervals: $[0,20),[20,40),[40,60),[60,150)$, already used in the delay analysis used in section 4.5.
2. Estimate $\overline{v}(\bullet)$ from the simulation results by taking the average delay for each of those VDS intervals.
3. Analogously, obtain the average prices p_{PEXIC} from simulation results.
4. Calculate p_{BIC} by calculating the Simpson Integral on $[EP]$ with the average delay values found in point 1. The resulting free constant term is obtained by imposing strict budget balance on p_{BIC} by weighting the curve with ϕ, the population density of type.

How does the BIC procedure translate to an actual mechanism? Since we are comparing asymptotic equilibria values, the former BIC expression would correspond to a mechanism in which PEXIC's phasing allocation rule is solved as in the original work but the calculation of prices is postponed at the end of the period of study or every time (if we want it to be strictly budget balanced). Alternatively, the average allocation can be learned and the average prices charged directly from the resulting learned expression. A set of final rebates would lead the mechanism to be approximatively budget balanced.

The next figure shows a comparison of both prices. It can be seen that the average price differences between PEXIC and BIC are very little (generally less than \$0.01). Assuming that agents are expected utility maximizers, we conclude that PEXIC is asymptotically Bayesian-Incentive-Compatible.

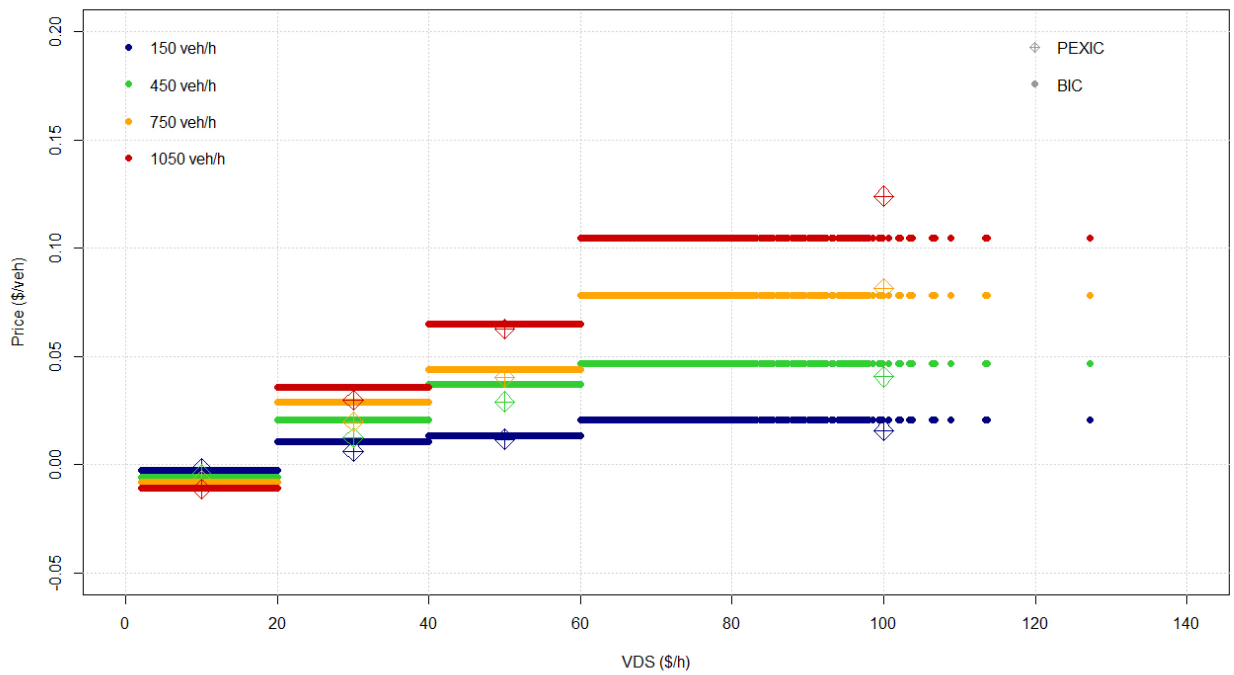


Figure 4.11: Comparison of minimum envy prices (PEXIC) vs. Bayesian Incentive Compatible (BIC) prices

Potential (polynomial) specifications are the only family of functions whose integral has the same functional specification as the original function multiplied by the base variable. This is useful in which the functional specification for $\bar{t}_i(\theta_i)$ also respects $[EP]$ in its continuous form. In the future, further exploration is possible for this specification, as well as pseudo-logarithmic specifications, which also share the same algebraic property as polynomial functions.

Finally, some general thoughts can be expressed on online incentive compatibility. In contrast to applications where online incentive-compatibility has problems, such as customers plugging in the EV vehicles for charging (Enrico et al., 2011) or customers joining Wi-Fi networks (Friedman and Parkes, 2003), vehicles under PEXIC traffic control system are permanently detected in the system due to full connectivity, therefore it is not possible for them to lie to the system on whether they will appear or not. Furthermore, this also complicates misrepresenting the arrival time to the intersection by accelerating or decelerating on purpose, since trajectories are monitored. But privacy can be preserved, since this layer of information (trajectory) is only shared in an encrypted way with each local controller.

4.8 Conclusion

This chapter presented a new method to optimize the operation of intersections differently than in the current practice, in that it considers the heterogeneity in value of delay savings of the individual users. The method presumes that current technology renders data on such individual valuations available for control. We proposed a mechanism that allows for signal timings to be determined without insisting on the traditional first-come first-served (FCFS) order, which we believe can be broken if fairness in service can result from utility exchanges via payments among users. We also demonstrate that such a system can be more efficient.

The proposed intersection control policy successfully manages to increase efficiency and fairness while including vehicles' value of delay savings for a large range of critical volumes and users' heterogeneity. We introduce a novel fair pricing approach which compensates for the additional delay differences between agents by means of utility transfers. Fairness is addressed by minimizing the envy perceived by users about the monetary transfers and experienced delay of others. Furthermore, the control mechanism is budget-balanced, and therefore it does not require any external subsidy to operate. The mechanism is modeled on a commercial microsimulation software and conclusions are drawn on the resulting asymptotic average indicator values using a variable sensitivity analysis.

In contrast with current approaches that minimize total delay, this scheme minimizes the total cost of all vehicles (cost being each individuals' value for their delay), by accounting for the value heterogeneity across the vehicles. Cost savings found are up to 20%, and these savings decrease as the volume level increases. As it is the costs that are minimized, the delays could be higher than in the delay-minimization case. We however find an increase of only around 15%, and it is consistent for across different volume levels. Finally, the resulting combination of travel cost reduction and fair compensation produces average utility increases for the entire value of delay savings range. This means that nobody is worse off under this scheme in comparison to delay minimization. We achieve Pareto efficiency in the form of travel cost reduction by including individuals' value of delay savings in a state of the art looking-ahead adaptive traffic control algorithm. We have proved that the inclusion of the multi-agent component increases the computational complexity of the phase allocation algorithm, but it still remains polynomial, which makes the control usable for real-world conditions.

A feature of the study is that user's values for delay savings are drawn from a semi-empirical distribution that takes advantage of the latest findings on value of time, scheduling delay savings and value of urgency. This very last feature is especially relevant, since it is observed to be the main reason which explains willingness to pay for travel cost savings. We qualify our distribution as conservative in the sense that we did not include abnormally high valuations that could always justify the policy. Our objective was to explore the consequences of this novel approach on a future implementation. The control is naturally sensitive to valuation heterogeneity, but achieves efficiency increases even under distributions close to well-known distributions of value of time savings estimated with field observations that do not show extreme heterogeneity.

Perhaps the most important feature of the presented scheme is the fair pricing scheme that results from an envy-minimizing optimization program, an approach imported from normative welfare economics. To our knowledge, it is for the first time that this methodology is introduced into transportation literature. This concept addresses fairness by compensating utility differences from valuation discrimination on delay. It is informationally not demanding in the sense that the comparisons do not require users to know others' utility specifications (interpersonal utility comparisons). The compensation program is linear (LP) and is easily solved by standard methods.

PEXIC's results, properties and operational flexibility show a great potential for generalized priority intersection control, a particular case of peer-to-peer exchange of supply in transportation. Nevertheless, many issues need to be addressed. Firstly, further empirical studies on user behavior and savings valuation need to be done. For instance, it would be worth studying the role of risk aversion and reference-dependent utilities or other notions such as the valuation of impatience. This could also lead to studying the effect of the valuations changing as a result of the experienced travel times, as each driver proceeds through intersections on his/her route. Operation of this kind of systems with mixed fleet of vehicles, some with drivers and some driverless, where the operators of the driverless vehicles may have their own pricing mechanisms, brings up another dimension for fruitful research.

Secondly, the policy needs to be evaluated on network-level simulations to assess the efficiency scalability. Users exchanging supply in real time may induce changes in OD patterns and individual budget imbalances. An interesting possibility is the directional effects in urban networks, such as when drivers with higher urgency may be driving in the radial directions to CBDs, while drivers of lower urgency may be more prevalent in directions orthogonal to them. This also raises possibilities of flow control on routes on the basis of subsidies (or credits) that would channel flows towards more efficient use of network capacity.

Finally, it is worth further exploring until which point users can manipulate on real-time the outcome of the control for their personal benefit and how to prevent such selfish manipulations. Research on these aspects would determine the operational feasibility of this idea as well as its eventual public acceptance. Of no lesser significance are the studies required on the regulatory regimes to be put in place for such systems. There are indeed socio-political questions that may naturally arise in such schemes that involve payments across travelers, and thus policy questions are also a rich field to be examined. It suffices to say that this chapter only presented the first exposition of a method whose eventual application and success will require much further studies, both model-based as well as implementation-based.

Chapter 5

Traffic flow and parallel queue operations

This chapter explores peer-to-peer consumption and exchange of supply on traffic operations in two applications:

1. Dynamic cooperative trading queue routing control scheme for freeways and facilities with parallel queues.
2. Queue-jumping operations for exiting freeways.

The first application is designed to operate on a connected vehicle environment. The second applications requires a semi-autonomous connected environment.

5.1 Dynamic queue routing control scheme with cooperative trading for freeways and facilities with parallel queues

Connected vehicle environments bring new opportunities in the operation of traffic infrastructures. Until now, vehicles traveling on a link selfishly self-organized themselves into a service order which corresponds on average to a First-Come First-Served (FCFS) order, without any mutual exchange of urgency information. Lane changes naturally violate the FCFS service order through a distributed mechanism for gap acceptance, with very limited or no precise future delay information. That is, any violation of FCFS in current systems happens with no coordinated mechanism, whereas it is conceivable that such FCFS violations can be efficient. New connectivity can increase the efficiency of lane changes by offering better future delay estimation and and allowing the incorporation of a new variable: travelers' Value Of Time (VOT).

If it was possible for users to be informed of the downstream traffic conditions for each lane, i.e. the downstream delay on each lane at a particular timestep, and if they could communicate their VOT values to each other, travelers could decide which lane changes are the most beneficial for everybody. Thus, agents can violate the initial order by creating coalitions which give incentives to agents in front to choose longer queues in exchange of a side payment. This would lead to a different level of service for each lane or queue,the less congested lanes or queues then becoming faster than the more congested ones and preferable for the higher-VOT travellers who may be willing to pay. The concept presented in this section is applicable to both facilities with parallel queues (multiple parallel servers) i.e. access gates at ports, traffic intersections, and bottlenecks in a freeway section. We will use the terms "queue" and "lane" indistinctively, as the essential level of performance of each lane is captured by the queue associated with it, for the conceptual purposes of this section.

This section presents a new dynamic queue routing control scheme which violates FCFS and outperforms it in efficiency while being core-stable. The policy is now outlined. Agents can choose which queue or lane they want to switch to. Naturally, vehicles in front get to choose earlier. Agents are assumed to have perfect knowledge of the delay per lane (or queue) as well as the delay increases due to nearby vehicles' lane changes. Agents can communicate their values of time to any other vehicle they want to interact with. Agents can form coalitions and exchange payments among them to improve their utility. Our mechanism implements the most efficient allocation and ensures that all agents and any subset of agents present in the system are better off by participating and cooperating with the outcome solution. This is ensured by employing the concept of the core, the pillar of cooperative game theory. Broadly, the core is a feasible set defined by constraints which define the stability of coalitions based on their worth. We hasten to add here that such exchanges may not be legally allowed in traffic systems; however, we assume that demonstrated social/system efficiency can lead to regulatory changes in future.

Contrary to most common applications in cooperative game theory, traffic operations present externalities. This means that the worth of a particular coalition depends on what other coalitions do. This brings us to the domain of partition function games (PFG), a superset of the more commonly used characteristic function games (CFG), which are not complex enough to express externalities. Equivalent stability concepts to the core are defined for partition function games. In particular, we are going to use the concept of a strong core (Chander, 2014), which we believe is small enough to be meaningful and apparently non-empty for the simplified version of the current application. A fundamental result in Ichiishi (1981) establishes an equivalence between strategic games and partition games. This section will actually relate both approaches, since the strategic-cooperative interaction is modeled as the optimization of the union of n-level Stackelberg games with coalitions.

We claim the following contributions in this section. First, we present this novel operational scheme for parallel queues and freeway management. Second, we are the first to use and solve a multiple-discrete-strategies n-level Stackelberg game with coalitions. Third, we found that the problem appears to be always strong-core stable for the vertical queue case, and that it is generally stable for the horizontal queue case as well. Finally, we propose a new relaxation for the strong-core concept found in Chander (2014) and generalize it to the dynamic domain.

The section is organized as follows: section 2 presents the meaningful literature from microeconomics on queue games and the utilized strategic structures, section 3 presents a static version on the cooperative queuing problem, modeling the queue routing as a parallel static vertical queue, section 4 presents a dynamic version of the problem, modeling the queue routing as parallel horizontal queues, and section 5 presents the conclusions and further research.

5.1.1 Literature review

Microeconomics literature has extensively explored the stability, fairness and truthfulness of priority queues (Chun, 2016) in the context of single and parallel queues (Chun and Jeong Heo, 2008). In priority queues, an unordered set of agents with heterogeneous values of time occupies positions on a line valued with linear delay. The efficient queue ordering is the one which places the agents sorted by decreasing value of time. However, queues in transportation systems involve agents with physical dimension and not all queue orderings are possible due to agent obstruction.

Bradford (1996) studied pricing and incentives for a multiserver queuing facility. He analyzes both social welfare and operator's revenue but does not enforce budget balancedness nor cooperation of any kind. His results are based on steady states and reach standard marginal pricing conclusions in efficiency maximization.

The cooperative interaction between agents for this problem is represented by partition function games (PFG). PFG are normally classified by the externality, either positive or negative, that the formng of two coalitions creates on a third one. Hafalir (2007) explores the role of convexity on efficiency and core stability. Abe (2016) explores PFG with either positive of negative externalities. Similarly to characteristic function games, although less studied, several stability concepts have been developed for this more expressive concept (Hart and Kurz, 1983). These concepts being generally too large when non-empty, Chander and Tulkens (1995, 1997) propose the gamma core and the strong core Chander (2014).

The strategic interaction between travelers has an inherent arrival ordering. The most adequate equilibrium concept is that of a (multilevel) Stackelberg equilibrium. Much has been said about two and three level variants of this concept. Little has been studied on the more general multilevel case, however. Bialas and Chew (1982) explore coalition formation in multi-level Stackelberg games for linear resource problems.

On a related application, presented in section 5.2, a queue jumping mechanism is proposed for general purpose freeway operations, in which vehicles coming from upstream can pay queued vehicles for being overtaken. Stability in the problem emanates from envy-freeness, naturally found in position environments (Varian, 2007).

5.1.2 Parallel vertical queues: Static problem

Let us consider a section of road which has a downstream bottleneck of constant outflow. This bottleneck can either be a congested point or section of a highway, a saturated intersection, or in a more general sense, a multiserver queue at a port terminal or border crossing point. The downstream section has m lanes, and the upstream section has l lanes. m may be larger, equal or smaller than l. From each lane $l \in L$, a subset $M_l \subseteq M$ is accessible. There is a set N of vehicles approaching the queue from the back. Each lane $l \in L$ has $N_l \subseteq N$ vehicles. Each downstream lane m has a queue $Q_m \geq 0$ built up. Without loss of generality, we assume that $Q_{m'} \leq Q_m \forall m' > m$. These queues can represent actual queues of stopped vehicles or congested traffic at cell in a traffic flow model such as a link transmission model. For the analysis in this section, the facility dispatches one vehicle per unit of time per queue.

We decompose the ordered set of agents $N \backslash \{i\}$ in two sets. Let $A(i) = \{k \in N \mid k < i\}$ be the set of predecessors of agent i (i.e, an agent who arrived ahead) and $F(i) = \{k \in N \mid k > i\}$ the set of followers of i. Let $A(m,i)$ be the set of predecessors of i which choose lane m. Let $j_m = |A(m,i)| \; \forall m \in M$. This defines a lane choice set $\sigma : N \rightarrow M^N$. Contrary to priority queues, in our model, a traveler cannot advance past a predecessor unless he joins a queue which is shorter than the queue that the predecessor has joined.

The delay for agent $i \in N$ joining lane m, given a lane choice set $\sigma(N)$, is $d_i(Q, \sigma(A(i))) = (Q_m + j_m - 1)$ and the valuation experienced by agent i, $v_i(Q, \sigma(A(i))) = -\theta_i d_i(Q, \sigma(A(i)))$, where θ_i represents the value of delay in monetary units per unit of time. This variable will also be called the type of agent i. Agent i is charged a price p_i for bearing the delay $d_i(Q, \sigma(A(i)))$. Finally, his utility is $u_i = v_i(Q, \sigma(A(i))) - p_i$.

Upstream vehicles belong to different lanes and are ordered based on their proximity to the downstream boundary. Again without loss of generality, we disregard the lanes

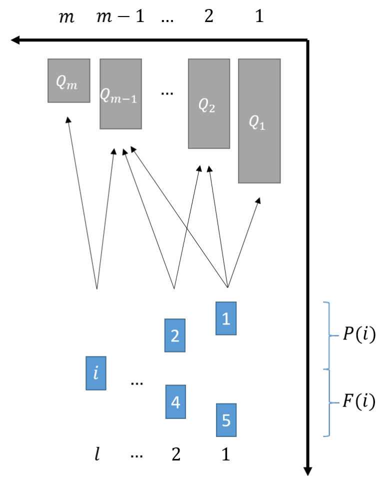


Figure 5.1: General problem configuration

l where vehicles are located, which is equivalent to assuming that $M_l = M$. If there is no communication between agents, vehicles will join the downstream bottleneck on a First-Come First-Served (FCFS) basis, each vehicle selecting the shortest queue. It is easy to see that if the initial arrival order is not monotonically decreasing on types, the resulting queue ordering will be inefficient, if we view efficiency in a utilitarian social welfare sense. However, if vehicles were to cooperate with each other, that is, forming coalitions to alter this initial ordering, a more efficient ordering for everyone would be achieved. This cooperation would be in terms of multilateral agreements on which lane every vehicle of the coalition would choose. This defines a multilevel Stackelberg game with coalitions. Of course, such coalition-forming would require communications and decision-making of the kind human drivers in traffic may not accomplish, but apps representing them can accomplish, and the technology for it certainly exists already.

Bialas and Chew (1982); Bialas (1989) explored some sufficient properties for stability for this kind of games, but on a linear resource allocation environment. We apply a similar recursion to our problem, but this time for pure strategies in extensive game. This recursion will give us the value generated by each coalition, given a particular coalition structure (partition of N). The recursion starts from the last vehicle and goes up. At

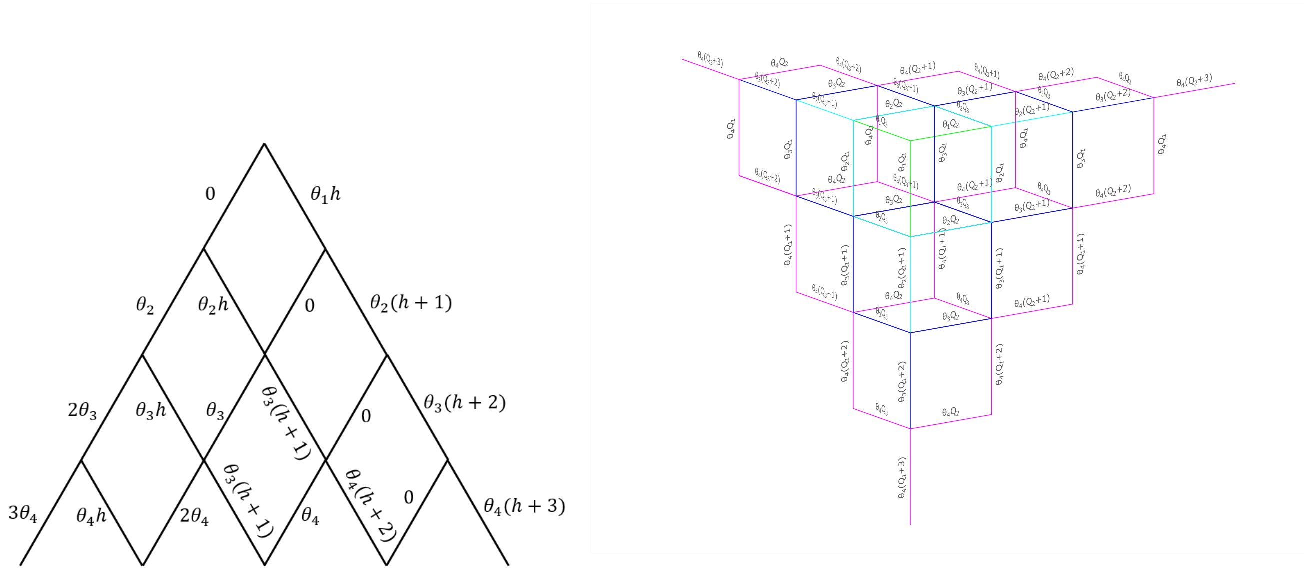


Figure 5.2: (left) 4-level Stackelberg with coalitions with 2 lanes (right) n-level Stackelberg strategy graph for 3 lanes and 4 agents

each level, the user i attempts to optimize the sum of the valuation of users who are behind it, $j \in N \mid j > i$ and which belong to the particular coalition i belongs to. The back recursion to solve the n-level Stackelberg problem with coalitions is, given a partition $P = \{S_1, \dots, S_j\} \mid \cap_j S_j = \emptyset$:

$$V^*_{S(i)|P}(h) = \max_{k \in K_i(h)} \left\{ v_i(h,k) + V^*_{S(i)|P}(h \cup k) \right\} \; \forall i = n-1, \dots, 1, \; h \in H_i \tag{5.1}$$

$$V_{S(n)|P}(h) = \max_{k \in K_n(h)} \{ v_n(h,k) \} \tag{5.2}$$

Where $S(i)$ is the coalition where i belongs.

(5.1) shows that for every user i, for every past history $h \in H_i$ up to user i, the agent selects the action k belonging to the set of available actions given h, $K_i(h)$, which maximizes the sum of two terms. The optimal value $V^*_{S(i)}(h \cup k)$ which results from the previous step $i+1$ and, $v_i(h,k)$, the valuation of user i from choosing action k, given history h.

Proposition 5.1. *The computational complexity of the recursion for the vertical queue case is $O(n(n+l)^l)$:*

Counting the number of nodes at the final level is equivalent to finding the number of l-combinations with repetitions:

$$\left(\!\!\left(\begin{array}{c} l \\ n \end{array}\right)\!\!\right) = \left(\begin{array}{c} l+n-1 \\ n \end{array}\right) = \frac{\prod_{i=1}^{l}(n+i)}{l-1} < \frac{(n+l)^l}{l-1} < (n+l)^l$$

Since there are n levels, the final complexity is $O(n(n+l)^l)$.

The recursion above needs to be solved for every partition $P \in \mathcal{P}$ to obtain all coalition values. The interaction between all the n-level Stackelberg games with coalitions will be modeled as a cooperative game. Cooperative games are complete-information games in which users are allowed to form coalitions to improve their payoffs. In the absence of coalitional externalities, cooperative games can be represented in a Characteristic Function Form (CFF), which defines a Characteristic Function Game (CFG). However,there are externalities in the current problem, which translates to coalitional payoffs being dependent on which other coalitions are formed. Thus, we will represent the game on a partition function form (PFF) which defines a Partition Function Game (PFG). Ichiishi (1981) shows that any strategic game can be represented as a partition function game. Thus, we will translate the former strategic games as a single PFG.

Let $P = \{S_1, \ldots, S_k\} \in \mathcal{P}$ be a partition of N such that $S_i \cap S_j = \emptyset \; \forall i \neq j$. We define the partition function$v: \; 2^N \times \mathcal{P} \rightarrow \Re$. That is, $v(S,P)$ represents the value of coalition S when the partition P is formed. $v(S,P)$ is in fact the sum of all the valuations $v_i \; \forall i \in S, \forall S \in P$, coming from the optimal order resulting from the cooperation between agents belonging to S when the partition formed is P. The pair$< N, v >$defines a partition function game. There is a coalition that is of special interest, the grand coalition $S_G = N$, which is composed by all members of the participant set.

A fundamental question in cooperative game theory is if this total cooperation will happen. This is desirable when the grand coalition is the most efficient coalition. When the partition function game arises from a strategic game, the grand coalition is always efficient since the set of strategies of the grand coalition game includes all the strategies available in the other subgames. In fact, the grand coalition payoff is the shortest path on the graph defined by the recursion (5.1) when $P = N$. Conversely, a coalition can unanimously dissolve itself into singletons, since every member can still choose the same set of strategies.

A fundamental characteristic of PFGs are the externalities that form on a coalition by third coalitions merging or splitting. Literature defines two types of externalities, positive and negative, which we define next.

Definition 5.1. *Positive (negative) externalities:*

$$\forall C,S,T \mid C \cap S \cap T = \emptyset \text{ and } \forall \rho \in P(N-(S \cup T \cup C)): \ v(C;\{S \cup T, C\} \cup \rho) > (<) \ v(C;\{S,T,C\} \cup \rho)$$

Basically, a PFG displays positive externalities when two coalitions S and T merge, increases the value of a third coalition C, for any complementary partitionρ. Conversely, the externalities are negative if the value of C is decreased. It is easy to see that this game has positive externalities.

Let $\rho \in P(N \backslash (S \cup T \cup C))$. Suppose all agents forming coalition C precede those of coalitions S and T. Then, the merging of S and T has no influence on $v(C,\{C, S \cup T, \rho\})$ $\forall \rho \in P(N \backslash (S \cup T \cup C))$ and the externality is zero. If agents forming C go after S and T, the externality can only be positive or zero since $S \cup T$ either causes some agents to join longer queues or keep their positions when no improvement is possible, reducing the cost of the members of C, which manage to advance some positions or none. However, if members of C are between those of S, T or $S \cup T$, negative externalities may occur. We provide an example next:

Example 5.1. *The static problem with instance* $N = \{13, 2, 14, 41\}$, $Q = \{4, 1\}$ *has negative externalties. Let* $S = \{1, 4\}$, $T = \{3\}$,$C = \{2\}$. *Then, by solving the game with the recursion algorithm, we reach the case of* $0 = v(C|\{S \cup T, C\}) < v(C|S, T, C) = 2$.

A useful property some PFG's exhibit is superadditivity, which is defined next:

Definition 5.2. *Superadditivity for PFG's:*

$$\begin{gathered} \forall S, T \subseteq N \mid S \cap T = \emptyset, \\ \forall \rho \in N - (S \cup T), v(S \cup T; \{S \cup T\} \cup \rho) \geq v(S; \{S, T\} \cup \rho) + v(T; \{S, T\} \cup \rho) \end{gathered} \tag{5.3}$$

Superadditivity means that if two coalitions S and T merge, their total payoff is larger than when unmerged. For the sake of exposition, the following counterexample shows that this game is not superadditive:

Example 5.2. $N = \{1, 9, 5, 33\}$ *and* $h = 3$. *If* $\rho = \{1, 3\}$, $S = \{2\}$, $T = \{4\}$,*then* $v(S \cup T, \{S \cup T, \rho\}) = 15$, $v(S, \{S, T\} \cup \rho) = 9$, $v(T, \{S, T\} \cup \rho) = 33$.

PFG's stability concepts and characterizations generally focus on games which have either positive or negative externalities Abe (2016) or exhibit superadditivity or convexity Hafalir (2007). This is not the case of our problem. We found an exception in the literature, which is the strong core for PFG Chander (2014), which is defined next:

Definition 5.3. *Strong-core for PFGs Chander (2014)*

$$\begin{gathered} (x_1, \ldots, x_n) \in \Re^n \mid \forall P \in \mathcal{P}, P = \{S_1, \ldots, S_p\} \neq [N], \\ \exists S_i \in P, |S_i| > 1 \mid \sum_{j \in S_i} x_j \geq v(S_i, P) \; and \; if \;\; P = [N], \; x_i \geq v(i; [N]) \; \forall i \in N \end{gathered} \tag{5.4}$$

The definition above states that for every partition which contains non-singleton coalitions, there exists at least one coalition of those coalitions which is worse off than in the strong-core imputation $(x_1, \ldots, x_n)$. Imputation is a term used in game theory to denote the utility agents obtain from a coalitional agreement. Moreover, every agent in the all-singleton partition is worse off. Contrary to other core specifications found in the literature such as the $\alpha, \beta, \gamma, \delta - cores$ (Hart and Kurz, 1983), the strong core does not assume any coalition structure for the complementary partition when a given coalition forms. The solution concept is particularly useful for the treatment of the status-quo partition, which, in our case, corresponds to the FCFS queue allocation. If an imputation vector belongs to the strong core, then all the singleton coalitions belonging to the coalition made of just singletons are better off by merging into the grand coalition. Now the question that remains is to prove non-emptiness of the strong core. As shown above, our problem has both positive and negative externalities. In order for the strong core to be non-empty, the PFG needs to satisfy two conditions:

Theorem 5.1. *(Chander (2014)) A PFG with general externalities $< N, v >$has a non-empty strong core if :*

1. v is partially superadditive: $\forall P = \{S_1, \ldots, S_m\} \in \mathcal{P}, \mid |S_i| > 1 \ \forall i = 1, \ldots, k \ |S_j| \ \forall j = k/+1, \ldots, m \ k \leq m, \ \sum_i^k v(S_i, P) \leq v(S, P') \ P' = P \backslash \{S_1, \ldots, S_k\} \cup \{\cup_{i=1}^k S_i\}$

2. and $< N, w^\gamma >$is balanced, where $w^\gamma(S) = v(S, \{S, [N \backslash S]\}), S \subset N$.

Partial superadditivity is weaker than superadditivity, which the game does not satisfy. Partial superadditivity is trivially satisfied for games with 3 or 4 players whenever the grand coalition is efficient.

Conjecture 5.1. *The strong core for the static queuing game as vertical queue is non-empty.*

We leave the result as a conjecture since it was not possible to prove it. After having run a large number of simulations, we did not find any counterexample, either. The values employed for the simulation study are: $n \sim Unif(1,\bar{n})$, $\overline{n} \in [2,7]$, $m \in [1,4]$, $\theta \sim logn(\mu = 2.16, \sigma = 0.7)$, $Q_m \sim Unif(1,4)$, $Q_i \sim Unif(1,Q_{i+1})\ \forall i < m$. The semi-empirical distribution used to draw individual Valuation of Delay Savings is developed in Chapter 4.

Standard proof methods, such as direct proof used in operations research for Minimum Spanning Tree Games and Shortest Path games Borm et al. (2001), did not prove successful. Neither did other approaches such as reduction to market games. Moreover, the game did not prove to be convex, which is a sufficient condition for non-emptiness of the strong core.

Instead, we are going to evaluate the inclusion of two imputations which are generalizations of the Shapley value for partition function games. These imputations satisfy the four properties of the original Shapley value: efficiency, symmetry, additivity and null player. The two imputations are presented next:

Definition 5.4. *(Do and Norde (2007); de Clippel and Serrano (2008)) Externality-free value:*

$$\phi_i^{free}(v) = \sum_{S \subseteq N} \zeta_S^i v(S, \{S\} \cup \{\{j\} \mid j \in N \setminus S\})\ \forall i \in N \tag{5.5}$$

Where:

$$\zeta_S^i = \begin{cases} \frac{(|S|-1)!(|N|-|S|)!}{|N|!} & if\ i \in S \\ -\frac{|S|!(|N|-|S|-1)!}{|N|!} & if\ i \notin S \end{cases} \tag{5.6}$$

The ζ_S^i values arise from the reordering of the marginal increment $v(S \cup \{i\}) - v(S)\ \forall S \subseteq N \backslash \{i\}$ expression, often found in the Shapley value definition, in terms of all the partitions $v(S)\ \forall S \subseteq N$.

This value represents that an agent leaving the grand coalition always creates a new coalition, that is, a singleton. This value is in line with the γ core present in the strong core existence theorem and we expect imputations from this value to generally belong to the strong core. The second imputation to test is:

Definition 5.5. *(McQuillin (2009)) McQuillin value:*

$$\phi_i^{McQ}(v) = \sum_{S \subseteq N} \zeta_S^i v(S, \{S, N \backslash S\})\ \forall i \in N \tag{5.7}$$

This value entails that an agent always chooses an existing coalition.

The following MILP will be used to test the feasibility of $\phi^{free}(v)$ and $\phi^{McQ}(v)$ in the strong core. We add the following objective function and modify the group rationality condition for the coalitions which have non-singleton partitions $P \in \hat{\mathcal{P}} = \mathcal{P} \backslash \{[N] \cup \{N\}\}$. We call this relaxation theϵ-strong core for PFG's, in line with the ϵ-core for characteristic function form games (Shapley and Shubik, 1966).

$$\min \epsilon \tag{5.8}$$

s.t.

$$\sum_{i \in S_j} x_i \geq v(S_j, P) - \epsilon - M z_{jp} \; \forall S_j \in \ddot{P} \subseteq P, \forall P \in \hat{\mathcal{P}} \tag{5.9}$$

$$x_i \geq v(\{i\}, [N]) - \epsilon \; \forall i \in N \tag{5.10}$$

$$\sum_{i \in N} x_i = v(N, \{N\}) \tag{5.11}$$

$$\sum_{S_j \in \ddot{P}} z_{jp} \leq |\ddot{P}| - 1 \; \forall \ddot{P} \subseteq P \in \hat{\mathcal{P}} \tag{5.12}$$

$$\epsilon \geq 0, \; z_{jp} \in \{0, 1\} \; \forall S_j \in \ddot{P} \subseteq P, \forall P \in \hat{\mathcal{P}} \tag{5.13}$$

Where $\ddot{P} \cup \dot{P} = P \mid \ddot{P} \cap \dot{P} \neq \emptyset$ are the collections of non-singleton coalitions $\ddot{P}$ and singleton coalitions $\dot{P}$ of every partition P. Essentially, the program consists of the relaxed group rationality constraints (5.9), the individual rationality constraints (5.10), and the grand coalition efficiency (5.11). The ϵ variable is the minimal slack for the most constrained coalition $S_j \in \ddot{P} \subseteq P$ necessary to make the problem feasible. The binary terms z_{jp} present in (5.9) and (5.12) enforce that at least one non-singleton coalition $S_j \in \ddot{P} \subseteq P$ for every $P \in \hat{\mathcal{P}}$ to be group rational.

The settings of this simulation are the same as for the strong-core non-emptiness evaluation. For each of these settings, 250 experiments are run. The next tables show the percentage of instances where the imputation was found in the strong core:

We observe that both imputations provide very similar percentages of inclusion into the strong core, the only differences being due to numerical errors from the solution algo-

$\phi^{free}: \overline{n}\backslash M$	2	3	4
2	100.0	100.0	100.0
3	100.0	97.2	98.0
4	96.0	85.2	88.4
5	87.6	75.6	72.4
6	84.0	67.6	64.0
7	74.4	52.8	61.6

$\phi^{McQ}: \overline{n}\backslash M$	2	3	4
2	100.0	100.0	100.0
3	100.0	97.2	98.0
4	96.0	85.2	88.4
5	88.0	75.6	72.4
6	83.6	67.6	64.0
7	74.4	53.2	62.0

Table 5.1: Percentage of experiments whose imputations are in the strong core.

rithm. As expected, the larger the vehicle set and the larger the number of lanes, the more the instances of not belonging to the strong core. This result seems to be weaker when increasing the number of lanes than when increasing the maximum platoon size. Both imputations having identical percentages suggest that a certain degree of symmetry exists in the violation of the strong core in the coalition formation process, for both all-singleton blocking coalitions and all-maximum-cardinality complementary blocking coalitions. This can provide some insights for the non-emptiness proof of this vertical queue model.

5.1.3 Parallel horizontal queues: Dynamic problem

Queues in transportation systems have a dynamic nature: they continuously receive arrivals and send out departures. Moreover, these queues are "horizontal", that is, queue length has a physical dimension. Modeling such a dynamic process as a succession of static queuing problems may lead to strong inter-temporal inefficiencies, coalitional instability and violation of individual rationality. One of the solutions, employed in section 5.2 is to set up a reserve price which prevents low value vehicles from having too much present savings at the expense of further high value vehicles. Another alternative would be to add a terminal cost at the leaves of the static problem tree; however this would require a transfer of payment from former agents to new agents as well as knowledge of further arrivals and their values, which would complicate the dynamic budget-balancedness and the efficiency of the system. In this particular problem, we will stick to a control solution which is always budget-balanced and efficient, aiming an eventual V2V decentralized implementation.

We model again a link with M lanes, but this time with length λ. Vehicles enter the link upstream at a constant speed v_a. There is a downstream bottleneck with a constant outflow q_{out} which corresponds to a headway h_q, spacing s_q and speed $v_q < v_a$. Every time an event happens, vehicles communicate to each other their value of time and positions, including the vehicles at the back of the queue. An imputation which satisfies the minimal ϵ-strong core is found and vehicles are assigned their new lanes. To minimize excessive perturbations due to the lane changes, only vehicles whose incorporation to the back of the queue is imminent will execute the lane change. The rest of vehicles will only execute the lane change once their incorporation becomes imminent. Naturally, the target lane can later change if there are further stability optimizations being executed due to new events happening.

Vehicles participate in lane-changing optimizations only as long as their incorporation to the back of the queue is not imminent. Once a vehicle is queued, they do not participate in other cooperative lane changes and are supposed to stay in the queue. The exchange optimization is run at every instant when there is a significant event. We define an event as the arrival of a new vehicle to the link or within imminent proximity of a moving vehicle at the back of the queue, or any external unpredictable event which could be detected by any of the vehicles. Events happen at time instants t, called epochs.

With this in mind, and using Newell's simplified car-following model, any vehicle's i predicted delay at epoch t is the maximum quantity of two situations: arriving to the downstream bottleneck undelayed at free flow speed, or being queued behind its predecessor vehicle:

$$d_i^t(S,P) = \max\{t_{i-1,t}^{dep}(S,P) + \frac{s_q}{v_q}, a_i\} - a_i \tag{5.14}$$

Since vehicles may participate in multiple optimizations, the utility specification is composed of the predicted total cost at epoch t since its arrival to the system minus the price charged after the optimization executed during epoch t minus the accumulated price charged to vehicle i until t. Vehcile i's imputation from being in coalition S and partition P is:

$$x_i^t(S,P) = v_i^t(S,P) - p_i^t(S,P) - \pi_i^{t-1}\ \forall i \in I_t, \forall S \subseteq P, \forall P \in \hat{\mathcal{P}} \tag{5.15}$$

$$v_i^t(S,P) = -c_i^t(S,P) = -\theta_i d_i^t(S,P) \tag{5.16}$$

With the utilities being defined, the dynamic ϵ-strong core program is:

$$\min \epsilon^t \tag{5.17}$$

$s.t.$

$$\sum_{i \in S_j} x_i^t \geq v(S_j, P) - \epsilon - M z_{jp}\ \forall S_j \in \ddot{P} \subseteq P, \forall P \in \hat{\mathcal{P}} \tag{5.18}$$

$$x_i^t \geq v(\{i\}, [N^t]) - \epsilon_t\ \forall i \in N^t \tag{5.19}$$

$$\sum_{i \in N} x_i^t = v(N^t, \{N^t\}) \tag{5.20}$$

$$\sum_{S_j \in \ddot{P}} z_{jp} \leq |\ddot{P}| - 1\ \forall \ddot{P} \subseteq P \in \hat{\mathcal{P}} \tag{5.21}$$

$$\epsilon^t \geq 0,\ y_{jp} \in \{0,1\}\ \forall S_j \in \ddot{P} \subseteq P, \forall P \in \hat{\mathcal{P}} \tag{5.22}$$

The program is identical to that in the static case, and is included for the sake of completion.

Parallel horizontal queues cannot benefit from the polynomial structure employed in the previous section. This time, the vehicles' costs depend not only on the queue length in front of them, but also on the dispatching times of the downstream vehicles. Since the departure times depend on the actual sequence of queued vehicles, a particular state is now defined by the sequence of vehicles in front of them, and therefore the whole queuing tree needs to be explored. This increases the computational complexity of the problem. For this reason, we will limit the number of participant agents of every optimization to six. Any additional vehicles upstream will stay outside of the exchange and continue advancing through the link on an FCFS basis.

Proposition 5.2. *The computational complexity of the recursion for the horizontal queue case is $O(l^n)$:*

For the horizontal queue case, all the tree histories need to be explored. This defines an l-ary tree with n levels. The last level has $O(l^n)$ nodes.

The exchange of information in positions also serves to define which lane changes are possible and which ones are obstructed. In our current formulation, if some lane change is not possible at a particular epoch t, the cost of its equivalent branch can be set to ∞ and that strategy will not get explored. However, this is not implemented here.

Next we explore the core stability of the dynamic problem as horizontal queues. We run 6 one-hour simulations for each of the scenarios defined by: $\lambda = 200\ m$, $L \in \{2,3\}$, $q_{in} \in \{360, 540, 720\}\ veh/h/lane$, $q_{out} = 900\ veh/h/lane$, $\theta \sim logn(\mu = 2.16, \sigma = 0.7)$. Arrivals arise from a binomial distribution for ease of implementation, but the simulation is still event-based. Both time and distance are continuous as well. The simulation is coded in MATLAB and the MILP programs are solved with Gurobi 6.0.5.

The next table (left) shows the percentage of optimizations which are contained in the strong score. We observe that for equal vehicle inflow, increasing the number of lanes increases core stability. This can be explained as the incoming platoon getting split into more lanes and their queues and interactions being smaller. Furthermore, increasing the incoming flow seems to increase instability, mostly due to a natural increase of complexity in the strategic interaction between agents. The table on the right displays the average ratio between ϵ and the average vehicle cost per optimization. This ratio is useful for comparing the magnitude of the slack term, i.e. the amount of utility that has to be transferred to a blocking coalition, with respect to the average cost of the participating agent. In line with the previous analysis, for equal inflow, a larger number of lanes decreases the magnitude of the instability. However, increasing the inflow seems to decrease the ratio, probably due to a larger increase in average vehicle cost.

$q_{in}(veh/h)\backslash M$	2	3
360	91.5	99.6
540	88.5	95.1
720	78.7	93.2

$q_{in}(veh/h)\backslash M$	2	3
360	0.29	0.17
540	0.11	0.16
720	0.04	0.02

Table 5.2: (left) % of strong core stable optimizations. (right) ratio $\epsilon/av.cost$

5.2 Queue jumping operations

5.2.1 Introduction

One of the successes of envy-free pricing is Generalized Second Price (GSP) auction for internet advertising (Varian, 2007). GSP auctions belong to the family of position auctions in which an ordered sequence of positions needs to be allocated to agents that value them differently. We bring in this idea to a new practical problem: queue-jumping operations on highways. By taking advantage of the recent advances in connected and automated vehicles, we show how exchanging transportation supply among users by means of an envy-free pricing mechanism outperforms FCFS queue operations.

Suppose there is a highly dense lane that leads to a highway exit. This lane could be a main lane or an exit lane with limited capacity. Adjacent to this lane there could be one or more lanes which we assume to be less congested. Normally, vehicles would join the queue at its back end on an FCFS basis. However, efficiency could be increased by allowing highest valuation vehicles to "jump" some of the queued vehicles when possible. Figure 5.3 illustrates the problem setting.

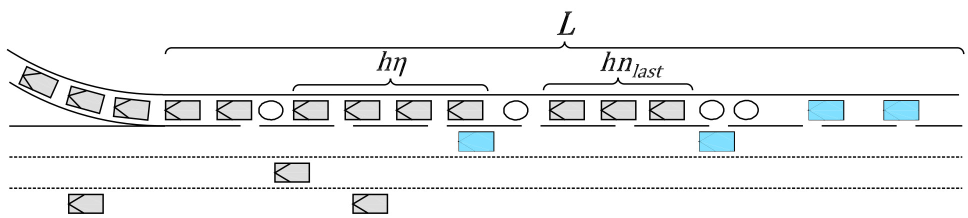


Figure 5.3: Queue scheme

5.2.2 Problem formulation

We assume that queued vehicles at the front of the queue travel at a speed v_q with an average spacing of s_q. A vehicle that enters the queue somewhere other than at the back end, which we call "jumper", has two effects in the system. First, it delays the overtaken queued vehicles, the "queuers", by $h = s_q/v_q$ seconds, potentially creating shockwaves that perturb the queue dynamics. We are going to assume that under a self-driving connected vehicle environment these shockwaves can be controlled such that the queue is always cleared at a rate of $1/h$ veh/s. Second, a vehicle taking a gap prevents a future vehicle from taking that same gap. We may not be interested in filling all the gaps at every auction if a particular time interval contains only low priority vehicles. Thus, we set a reserve price r to prevent these vehicles from taking those gaps and reserve them for further incoming high value vehicles. However, setting the reserve price too high may limit the efficiency gains: r sets an efficiency trade-off.

Our mechanism fixes a priori the supply of gaps, reserving a potential gap every η queued vehicles. We call each set of η vehicles between two gaps a "block". Gaps are labeled $M = 1,...,m$, where the first element is the first gap in the queue and m represents the back of the queue, with no delay savings. The last segment of the queue will normally be composed of $n_{last} < \eta$ vehicles. Ordering the supply in this manner controls for queue disturbances from "jumpers" and maintains overall queue efficiency. These gaps will not be physically reserved beforehand, but only created once a vehicle is assigned to them. A second design parameter is the minimum number of gaps μ needed for the policy to be activated. If there are less than μ gaps present, the queue is operated FCFS.

We next describe the queue-jumping policy, detailed in Algorithm 2. Suppose there is a residual queue Q_0. A vehicle i which desires to join the queue is detected with a speed v_a at a distance L from the front of the queue. A priority value b_i is elicited from vehicle i, i.e. a value of delay-savings (VDS) in $\$/s$. In the event that the minimum number of gaps condition remains satisfied, the mechanism will keep accepting arrivals such that any of them could be allocated to any of the available gaps. This is represented by the condition in line 9: every new vehicle has to be able to make it to the first gap before it departs or before the leading vehicle does not reach yet the back of the queue. These vehicles, together with leading vehicle i, will compound vehicle participant set N. The set of jumpers and queuers are noted as J and $N \setminus J$ respectively.

Algorithm 2 Queue-jumping mechanism

procedure QUEUE-JUMPING ALGORITHM
    $Q \leftarrow Q_0$
    $i \leftarrow detect\_arrival()$
    $M \leftarrow retrieve\_gaps(Q)$
    if $|M| \geq \mu$ **then**
        $t_{bq} = time\_to\_position(i, |Q|)$
        $\tau \leftarrow gap\_departure\_time(1)$
        $N = i$
        while $clock\_time() < min\{t_{bg}, \tau - L/v_a\}$ **do**
            $j \leftarrow detect\_arrival()$
            if $j \neq \emptyset$ **then**
                $N \leftarrow N \cup j$
        if $|N| > 1$ **then**
            $(\mathbf{x}^*, \mathbf{p}^*) = EFQOPT(N, M)$
            $\mathbf{p}_J \leftarrow \mathbf{p}_J^* + r\boldsymbol{\delta}_J(\mathbf{x}^*)$
            $\mathbf{p}_{N \setminus J} \leftarrow \mathbf{p}_{N \setminus J}^*$
            $\mathbf{p}_Q \leftarrow \mathbf{p}_Q - rh\mathbf{B}\mathbf{x}^*$
            $Q \leftarrow update\_gaps(\mathbf{x}^*, Q)$
        else if $b_i \geq r$ **then**
            $\mathbf{x}^* = (1, 0, ..., 0)$
            $\mathbf{p}_J \leftarrow r\boldsymbol{\delta}_{i1}$
            $\mathbf{p}_Q \leftarrow \mathbf{p}_Q - rh$
            $Q \leftarrow update\_gaps(\mathbf{x}^*, Q)$
        else
            $Q \leftarrow Q \cup i$
        goto 3
    else
        $Q \leftarrow Q \cup i$

When $|N| > 1$, the position auction EFQOPT is held and the optimal queue positions $\mathbf{x}^*$ and optimal prices $\mathbf{p}^*$ are calculated. Jumper vehicles are charged $p_J = p_J^* + r\boldsymbol{\delta}_J(\mathbf{x}^*)\ \forall i \in J$, where $\boldsymbol{\delta}_J(\mathbf{x}^*)$ is the vector of optimal delay savings. Queuers, on the other hand, get paid the optimal price $\mathbf{p}^*_{N\setminus J}$. Queued vehicles get paid for the externality suffered: $rh\mathbf{B}\mathbf{x}^*$, where $\mathbf{B}$ is an indicator matrix which shows which gaps affect which queued vehicles. When the participating set contains only the leading vehicle i, and its VDS b_i is greater or equal than r, this vehicle is allocated the first gap and the vehicle pays its delay savings times the reserve price: $p_J = r\boldsymbol{\delta}_i(\mathbf{x}^*)$. If $b_i < r$, the vehicle joins the queue at the back. The envy-free auction is modeled by the following mixed binary linear program, EFQOPT:

$$\max \sum_{i\in I, j<m} (b_i - r)\delta_{ij}x_{ij} \tag{5.23}$$

s.t.

$$\delta_{ij} = \max\{h((m-1-j)\eta + n_{last}), 0\}\ \forall i \in I, \forall j < m \tag{5.24}$$

$$(b_i - r)\sum_{k\in J}\delta_{ik}x_{ik} - p_i \geq (b_i - r)\sum_{k\in J}\delta_{jk}x_{jk} - p_j\ \forall i \in I, j \neq i \tag{5.25}$$

$$\sum_{j\in J} x_{ij} = 1\ \forall i \in I \tag{5.26}$$

$$\sum_{i\in I} x_{ij} \leq 1\ \forall j < m \tag{5.27}$$

$$(b_i - r)\sum_{k<m}\delta_{ik}x_{ik} - p_i \geq 0\ \forall i \in I \tag{5.28}$$

$$\sum_{i\in I} p_i = 0 \tag{5.29}$$

$$x_{ij} \in \{0,1\}\ \forall i \in I, j \in Q \tag{5.30}$$

Here, δ_{ij} is the delay savings of vehicle i taking gap j, defined in (5.24) and x_{ij} is a binary variable which takes the value of one when agent i is assigned position j, and zero otherwise. These savings are decreasingly ordered with j, with $\delta_{im} = 0$ applying to the vehicles that join the queue at its back on an FCFS basis. The objective function (5.23) is the aggregate welfare: efficiency increment due to queue skipping minus the externality created on the vehicles overtaken. The envy-free conditions (5.25) include the externality on overtaken queued vehicles directly as a price, taking r as an upper bound of the queued types. Note that jumpers are never overtaken after jumping, since they are allocated the very first available gaps due to social welfare maximization under linear utilities. The only scenario where a vehicle with type larger than r can suffer envy is when the demand is larger than the supply of gaps. Then, such vehicles get queued, and compensated as if it was of type r. This generally happens only once, since there will be only one gap available in front of such a vehicle, unless the demand is high enough that many batches are formed by FCFS vehicles. Learning and choosing proper η and r values is crucial to achieve maximum efficiency increases. (5.26) and (5.27) are the assignment constraints, which restrict one vehicle per position, except for the position corresponding to the back of the queue. (5.28) is the individual rationality constraint and (5.29) is the budget balancing constraint. Note that the payments to the queuers do not figure in (5.29) since they are fixed and cancel out anyway. An example of queue operation is shown in Figure 5.4. The time dimension shown is after a horizontal affine transformation which subtracts the time to travel a distance x at a speed v_a. This makes the trajectories vertical and it is useful to know the set of accessible gaps for each vehicle, directly from its arrival to the system. The vertical axis is divided by v_a to express the distance as seconds traveled at the approaching speed.

There is an existing queue formed on an FCFS basis. Suppose that μ was set to 3. A vehicle arrives and activates the mechanism since there are three gaps now. From that time now until the departing time of the first gap (shown by the first vertical blue line and the top dashed red line), all the comprising vehicles are selected for the first auction. We say that a gap departs, when its trajectory reaches the front of the queue. The supply

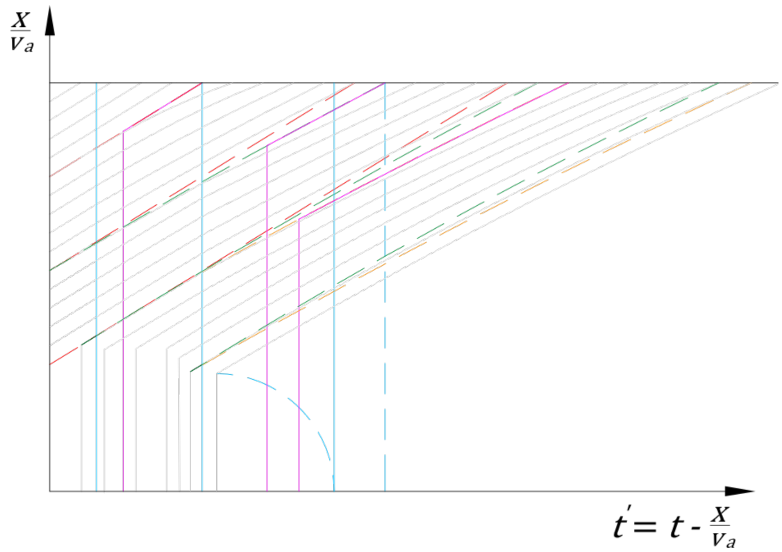


Figure 5.4: Trajectories. Red dashed lines: gap trajectories if FCFS. Purple lines: jumper vehicle trajectories. Blue vertical lines: auction range delimiters.

set for this auction are the first two gaps, leaving the last gap always for the next auction. The participant set for this auction, N_1 are all the vehicles between the leader's arrival time and the departure time of the first gap. The auction allocates the third vehicle the first gap, and the rest of the vehicles join the back of the queue. Notice how the gaps are only created once a vehicle has been assigned to them. Thus, unfilled gaps do not produce efficiency losses.

The second auction starts with the first arrival after the first gap has departed. Spare gaps from the first auction are now displaced one position back. Their trajectories are shown by the dashed green lines. This time, the participant set N_2 is constrained by the time the leading vehicle takes to reach the back of the queue, rather than the departure time of the next gap. This is shown by the blue dashed circular line and the second solid blue line. This auction allocates the second and third vehicles gaps 1 and 2, respectively. Finally, for the next auction, since $|M| < \mu$, any further incoming vehicles would join the queue on an FCFS basis. Trajectory of the last gap for a further third auction is shown in orange.

The model so far leaves out the dynamics and operational details of gap creation. This is justified since the economic gains depend only on the output rate at the front of the queue, the outgoing headway h. Here, h is considered constant, but the policy can be easily generalized to variable headways as long as h does not change drastically. Delay savings being dependent on the headway and the number of vehicles skipped makes the mechanism independent of how the gaps are created, leaving it open to flexible gap creation policies.

In a real implementation, the infrastructure manager can adjust η to minimize potential perturbations or limitations due to congestion from and towards the adjacent main lane. This is relevant for the gaps that need to be actually created from an already present queue, such as the first and second gaps in Fig. 5.4. However, when the gaps are at the back of the queue, these gaps can be created by just slowing down the queuers from the previous auction and letting the jumper vehicle overtake them. This would be the case of the second and third gaps in the second auction.

Proposition 5.3. *The queue jumping mechanism problem is feasible and has a unique optimal allocation x^* and multiple prices p^* that can support such an optimal allocation. Such a solution is Pareto optimal and envy-free, and the mechanism's efficiency is always equal or superior to FCFS.*

Proof. Suppose $b_i \neq b_j$, $\forall i \neq j$. Due to allocative efficiency maximization, the allocation is order monotonic: a vehicle i with $b_i > b_j$ will always have more advanced position than j. By Theorem 3.4, there is a range of envy-free prices that support x^*. By Theorem 3.1, the mechanism is also Pareto optimal. Finally, since the objective function can only take non-negative values and the externality on queued vehicles is compensated by individually rational payments from jumper vehicles, the mechanism's social welfare is always larger or equal to that of FCFS. □

It is informative to examine how this queue-jumping policy fits into the conceptual framework presented in Section 2, so as to both view the interactions properly and to evaluate the sufficiency of the Section 2 framework. The operator objective O is efficiency maximization, the environment E is envy-freeness and individual rationality, under honest bidding. Since the policy is budget balanced, operator's resources R is the empty set. The allocation and price rule $[T, F]$ is Algorithm 2 and the collaborative exchange procedure is the execution of Algorithm 2.

5.2.3 Simulation results

Finally, we simulate the mechanism over $T = 1800$ seconds and 6 random seeds. The ramp length L is 300 m. The types are randomly drawn from a log-normal distribution with μ =2.16 and $\sigma = 0.7$. The bottleneck flow rate is 833.33 veh/h which comes from a queue spacing s_q= 6 m and v_q= 5 km/h and the incoming flow rate follows a Poisson distribution with an average flow rate of 900 veh/h. Therefore we have oversaturated conditions which build a queue but we set an initial queue Q_0 of zero vehicles. Incoming vehicles travel at v_a= 80Km/h. We do a sensitivity analysis on $r \in \{5, 10, 15, 20\}$ \$/h and $\eta \in \{2, 3, 4\}$ veh/block. Figure 5.5 shows the resulting efficiency increase (objective function (5.23) minus externalities from jumpers to queued vehicles) with regard to those variables. As expected, efficiency gains follow a concave relationship with the reserve price: if r is too low, too many low VDS vehicles take many gaps. On the contrary, if r is too high, not many vehicles can take advantage of the policy. Concerning η, the efficiency gains decrease as η increases, since the supply decreases. In the presence of congestion, however, we should expect smaller gains in the lower η region.

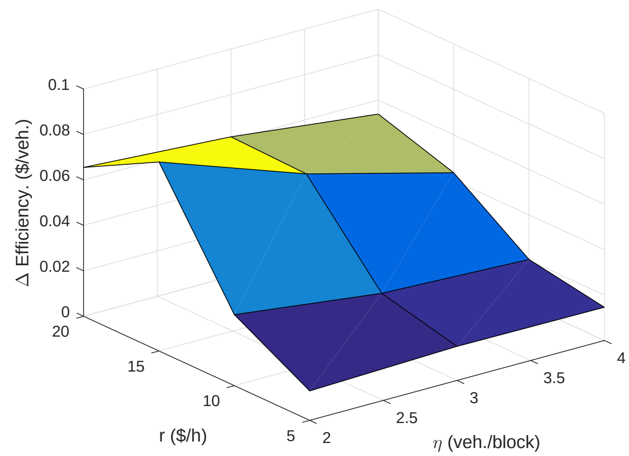


Figure 5.5: Efficiency increase in function of η and r.

5.3 Conclusion and further research

This chapter presented two applications of collaborative consumption of supply for flow traffic operations: cooperative queue routing for freeway operations and parallel queue facilities; and queue-jumping operations for exit lane in freeways. The first application concerns connected vehicle environments and the economic goal is core stability, while the second application requires a semi-autonomous connected vehicle environment and its goal is fairness and also queue stability through envy-freeness.

The first section presented a new collaborative control mechanism for freeways and parallel queue facilities. Under this control scheme, agents observe predicted future delays per lane (or queue) and are allowed to collaborate to change lanes such that the total travel cost of their platoon is minimized. High VOT vehicles can pay low VOT vehicles to switch to a more congested lane while they can stay in the same lane or switch to another lane with less vehicles in front. The underlying cooperative principle is the strong core for partition function games. While aimed as a decentralized, distributed control, this section assumes a centralized optimization to evaluate the economic efficiency and stability without excessive technicalities.

The control policy has been first explored as a simpler vertical queue model. In this case, the strategic interactions between users form a tree-like structure which is of polynomial-time complexity. While not proved, simulation results suggest that the problem may be strong-core stable. In addition, we have tested two generalizations of the Shapley value for partition function games which we found to generally be strong-core stable.

We then modeled the control policy as a dynamic horizontal queue. We have observed that the policy is generally strong-core stable except for situations when there are sharp increases in the incoming platoon size. However, it is computationally intensive. Further distributed optimization techniques should be used to make it applicable for an eventual real-world implementation. Alternatively, designing an approximation algorithm for the strong-core optimization would be useful.

For further research, we point out the following lines. We believe that developing a formal non-emptiness proof for the vertical case would represent a strong result in cooperative game theory. Also, from a theoretical point of view, exploring the dynamic vertical queue case would be interesting in the sense that dynamic applications in partition function games have never been explored outside of coalition formation in static settings. Concerning the dynamic horizontal queue, we will further explore unstable instances to better determine the source of instability and better understand the problem. Eventually, modeling such policy with a commercial microsimulation software, would provide insights on the efficiency increases of a real world situation, as well as including more realism and efficiency losses due to lane changing obstruction.

The second application concerns queue-jumping operations for exit lanes in highways. We show that this operational scheme is Pareto efficient and fairer than FCFS. The fairness metrics is based on the fact that this queue environment is envy-free. In this application, envy-freeness also serves as a stability concept. The policy is also individually rational. As a further research, a deeper study is required for further assessment of economic benefits, especially in the presence of congestion in the adjacent lane to the queue. Also, performance modeling under more realistic car following models need to be carried out before considering any real implementation. Finally, new participating set formation policies could be studied if we allow the envy-free condition to be relaxed.

Chapter 6

P2P Ride Exchange in Dynamic Ridesharing Systems

6.1 Introduction

Peer-to-peer (P2P) dynamic ridesharing is a shared mobility alternative in which peer drivers and riders (passengers) share the space in the drivers' personal vehicles. The term "P2P" implies that drivers are not hired by companies to transport passengers, but are rather using their personal vehicles to carry out their personal tasks, which makes them peers to riders. The term "dynamic" highlights the fact that customers can join the system at any point in time and do not have to book their trips in advance.

P2P ridesharing manages to eliminate vehicles from roads by getting people who are traveling in the same direction in the same vehicle. P2P ridesharing benefits drivers, riders, non-users, the transportation infrastructure, and the environment. Drivers receive monetary compensation for the service they provide while following their own daily schedules, and riders are charged less than other transportation alternatives, such as taxis. By reducing the number of traveling vehicles and hence congestion levels, the benefits of P2P ridesharing are extended to the entire community as well as the environment.

Contrary to traditional service businesses where servers belong to the business and their number is proportional to the demand for service, in P2P ridesharing servers are also customers. Therefore, it is important for the system operator to attract the right proportion of riders and drivers. Another feature of P2P ridesharing systems is that drivers typically have specific locations where they start and end their trips, and tight travel time windows to carry out their trips. This limits the level of spatiotemporal coverage of the network by each driver. Therefore, in order to serve a higher number of riders, a ridesharing system needs to increase the spatiotemporal coverage of the network by increasing the number of drivers. To motivate, attract and retain a high number of drivers, a high number of riders is necessary. Therefore, the number of customers in a P2P ridesharing system should pass a certain critical mass with a specific proportion of drivers to riders, in order for the system to be able to operate independently and without a need for outsourcing supply.

A ride-matching algorithm is the engine of a P2P ridesharing system, determining how drivers and riders should be paired. Except for very simple and non-efficient ridesharing systems (which we will discuss later), the ride-matching problems are computationally hard to solve. A good ride-matching method is one that can provide the highest number of matches in an attempt to engage the highest number of customers and bypass the critical mass of participants.

Customer experience is another factor that plays a role in the success of a P2P ridesharing system, especially during the initial phases of implementing the system. A customer (rider or driver) may give the system a chance by attempting to use the system a few times, but if he/she is not matched, there is a possibility that such a customer would never return to the system. Therefore, it is essential for a P2P ridesharing system to involve and retain as many customers as possible.

Customers in many transportation systems are served on a first-come, first-served (FCFS) or a similarly pre-ordered basis. For P2P ridesharing, in which customer retention is especially important, considering riders on an FCFS basis is an inefficient use of the very limited available resources (drivers). The FCFS rule, however, is the natural order of serving riders in a dynamic system, where riders announce their trips not long before departure. In addition, dropping the FCFS principle may lead to high solution times for the resulting matching problem, and is therefore not an appropriate implementation strategy for a dynamic real-time system.

In this chapter, we introduce what we call *P2P ride exchange*, a mechanism to improve the number of matches in an FCFS-based system. In a system where P2P ride exchange is implemented, riders will still be considered for service on an FCFS basis. Upon joining the system, a rider will be offered the best available itinerary, according to certain criteria which we will discuss later. However, if no match exists, the rider will be given the chance to buy a previously-matched rider's itinerary under specific circumstances. Purchasing an itinerary from a previously-matched rider is in fact reversing the FCFS rule. This exchange of rides is accompanied with an exchange of money through the system. Since the objective of the system from implementing the exchange mechanism is to increase the total number of matched riders, only riders for whom an alternative itinerary is available will receive a proposal to sell their current itineraries.

There are, admittedly, considerable regulatory obstacles to overcome for such P2P exchange or trade schemes to be used in transportation systems. The legal battles faced by ridesourcing firms are now well-known. Transportation supply being considered a public good, any breaking of the traditional FCFS operational paradigms also could face objections based on socio-political arguments of inequity across users. While important, such topics are considered beyond the scope of this chapter that focuses only on showing the performance potential of the proposed scheme.

6.2 Related Work

P2P ridesharing systems are a member of the family of shared-use mobility alternatives. There is an abundance of work in the literature on the benefits ridesharing systems offer in terms of reduced direct and indirect cost to the environment and the society (Chan and Shaheen, 2012; Morency, 2007; Heinrich, 2010; Kelly, 2007). Despite these benefits, ridesharing operators have been facing multiple challenges in running ridesharing systems as stand-alone businesses. Furuhata et al. (2013) conduct a thorough survey of different types of ridesharing systems, and discuss some of the challenges that have prevented these systems from reaching their potential, despite the improvements in communication technology, prevalence of GPS-enabled cell-phones, and ease of developing cellphone applications that greatly facilitate participating in ridesharing systems.

Ultimately, for a ridesharing system to operate successfully, it has to attract and maintain a critical mass of customers. An essential challenge practitioners face is finding the most effective way to build this critical mass (Cervero and Griesenbeck, 1988; Brereton and Ghelawat, 2010; Raney, 2010). James Shield of Carma Technology Corporation which develops carpooling applications is quoted in an article by Gaynor (2015) to believe that although there is no definitive answer to this question, attracting a higher number of drivers, increasing marketing efforts, improving the technology, and attempting to use a societal/behavioral approach to engage people and make habits are all valid approaches.

Research in the field of marketing has found customer satisfaction, among other factors, to be a great predictor of customer retention rate (Gustafsson et al., 2005; Ranaweera and Prabhu, 2003; Rust and Zahorik, 1993). A satisfied customer not only has a higher probability of returning to the system, but also generates positive word of mouth (WOM) that helps in attracting new customers (Söderlund, 1998). Research has shown that WOM is a more important factor when it comes to deciding on services, rather than goods (Buttle, 1998). In addition, the cost of customer acquisition is about five times the cost of customer retention (Pfeifer, 2005), suggesting that a customer's first few experiences with the system play a central role in its long-term success. In light of these research studies, it is very important for ridesharing systems, especially in their initial stages, to serve as many ride requests at possible. In addition to a high matching rate, the responsiveness of the system to dynamic ride requests could play a role in customer satisfaction. Dynamic systems which try to address requests in real-time score high in this respect.

Encouraging a high number of drivers to participate in the system is another goal of a ridesharing system, albeit not as important as the first one. The reason is that firstly, drivers who participate in ridesharing usually receive a base fare regardless of the extent of their contribution. For a given level of demand (ride requests), there is an optimal amount of supply (drivers) above which the contribution of additional supply is only marginal, and therefore attracting drivers with only marginal contributions is not financially wise. Secondly, drivers in a ridesharing system are traveling to perform their personal activities. Even if they are not matched on a regular basis, entering their fixed daily schedules in the system only once could earn them extra revenue.

The ride-matching algorithm used by a ridesharing system plays an important role in the number of riders the system can serve. The simplest form of ride-matching algorithm matches each driver with a single rider (Agatz et al., 2011; Wang, 2013). This problem can be formulated as a maximum cardinality matching problem on a bipartite graph and solved quickly using efficient algorithms (Alt et al., 1991). The more sophisticated ride-matching problems are capable of allocating more than one rider to a single driver (Wolfler Calvo et al., 2004; Baldacci et al., 2004; Teodorović and Dell'Orco, 2005; Herbawi and Weber., 2012; Di Febbraro et al., 2013), proposing multi-hop itineraries to a single rider, where the rider can transfer between multiple drivers (Agatz et al., 2010; Masoud and Jayakrishnan, 2015*a*; Ghoseiri, 2013; Herbawi and Weber., 2011*a*,*b*), and finally considering multiple riders and drivers in the same problem, and proposing multi-hop itineraries at the same time (Masoud and Jayakrishnan, 2015*a*; Ghoseiri, 2013; Regue et al., 2016).

A many-to-many ride-matching problem, where a rider can transfer between multiple drivers and a driver can carry multiple passengers at any moment in time, is the most comprehensive form of ride-matching and can yield the highest number of matches. Not surprisingly, a many-to-many problem is also the hardest matching problem to solve. For a ridesharing system to enjoy the benefits of many-to-many ride-matching, it should have access to information on future trips of riders and drivers. This property of a many-to-many problem coupled with its higher solution time prohibit such a system from being used in a dynamic setting where matches need to be made in real-time and information on future trips is not typically available. Many-to-many ride-matching is, however, very effective in static implementations of ridesharing, where system participants are required to announce their trips by a certain deadline.

In a many-to-one matching problem, a rider can transfer between drivers, but the matching problem is solved for one rider at a time. By definition, a many-to-one problem provides the best (multi-hop) solutions for a dynamic system where riders need to be informed of the status of their requests as soon as they input their register their trips. There are two ways, however, to shift the solution of a many-to-one problem towards that of a many-to-many problem in a dynamic system. After a rider is matched in a real-time system, the itinerary of the rider is fixed, and the drivers constructing the itinerary will have to remain committed to their assignments. It is possible, however, to include previously matched drivers with fixed itineraries (that respect their previous assignments) in the matching problem for the current rider, increasing the level of supply available to the current rider and hence enhancing our chances of satisfying their request while at the same time increasing vehicle occupancies. Introducing this small variation will transform a many-to-one system in which each driver carries one rider at a time, to a many-to-many system, in which each driver can have multiple passengers on board. Note that the type of matching method used in this case is still a many-to-one method.

The second way of shifting the solution of a many-to-one problem to that of a many-to-many problem is by implementing the P2P ride exchange mechanism introduced in this chapter. Contrary to a many-to-many matching problem that assumes no particular order in serving requests and hence achieves a high matching rate as a result of introducing this additional level of flexibility, the nature of a dynamic system, which requires attending to requests as soon as they arrive, calls for an FCFS-based implementation. P2P ride exchange attempts to reverse the impact of the FCFS rule by proposing to a rider who has been offered an itinerary to switch to a less attractive itinerary, liberating the drivers contributing to the rider's current itinerary from their commitments. This exchange is motivated by a monetary compensation from a second rider who finds the liberated drivers more valuable. The P2P exchange mechanism proposes the amount of this compensation ensuring that the system will remain budget-balanced.

In this chapter in order to match riders and drivers, we use the many-to-one ride-matching algorithm proposed by Masoud and Jayakrishnan (2015*b*). We choose this algorithm because it can solve matching problems in real-time and can be easily modified for use in one-to-one and many-to-many ridesharing systems as well, giving us the ability to study the impact of the P2P ride exchange mechanism on a wide range of ridesharing systems. Furthermore, using this algorithm, not only can we find the optimal itinerary for a rider, but we can also identify and store other feasible itineraries that can be used later by the exchange mechanism.

There have been a few attempts in the literature to design mechanisms for para-transit and ridesharing systems. Furuhata et al. (2015) propose the Proportional Online Cost Sharing Mechanism for demand-responsive transport Systems. This mechanism is capable of proposing an upper bound on the fare a potential user has to pay. The mechanism relies heavily on having the passenger requests in advance of the start time of the trips. The focus of the work is on proving the online fairness, budget balancedness, individual rationality and ex-post incentive-compatible properties of the mechanism under certain conditions. The mechanism, however, unlike the P2P ride exchange mechanism proposed in this chapter, is not designed to increase operational efficiency of the system. Wang (2013) proposes a stable matching game between riders and drivers in a one-to-one system, where no rider/driver can be better of by unilaterally switching to other drivers/riders. Although such a system can lead to an equilibrium, for it to yield operationally efficient results, it requires access to the participants' trip information in advance. Kleiner et al. (2011) proposes an auction-based allocation mechanism that incorporates users' valuations on the ride assignment. While this mechanism violates the FCFS rule, it is not real-time since it uses a rolling horizon in which decisions are delayed.

P2P ride exchange is the first real-time mechanism designed to address the inherent trade-off between the two factors that influence customer satisfaction in a ridesharing system, namely rider matching rate and system responsiveness. To the best of our knowledge, the proposed P2P ride exchange mechanism is the first trading mechanism to increase ridership in a dynamic P2P ridesharing system. The designed mechanism is limited to bilateral trades, where there is a single buyer and a single seller. This mechanism, therefore, is optimal for a one-to-one matching system, and provides a lower-bound on the increase in ridership in one-to-many and many-to-many systems. In the rest of this chapter, we first provide a brief summary of the ride-matching algorithm used. We then officially introduce the mechanism, and elaborate on some of its properties. Finally, we conduct extensive numerical experiments to quantify the performance of the P2P ride exchange mechanism under different parameter values for the system.

6.3 One-to-many Ride-Matching Algorithm

We solve the ridesharing problem using the dynamic programming (DP) algorithm proposed in Masoud and Jayakrishnan (2015*b*). This algorithm is suitable for the purpose of P2P ride exchange because firstly, using this algorithm, real-life size problems can be solved in a very short period of time (a fraction of a second in most settings), and secondly, all feasible solutions to the problem are retrievable using the set of trees that are generated while solving the problem. In this section, we provide a brief review of the algorithm.

Let us define graph $G = (N, L)$. Each vertex $n \in N$ in this graph is a tuple $(s, t) \in S \times T$, where S is a pre-defined set of stations in the network where participants can start or end their trips and/or transfer between vehicles, and T is the set of time intervals during the study time horizon (set to be one minute in this study). An edge $\ell = (n_1, n_2) = (s_1, t_1, s_2, t_2) \in L$ in this graph corresponds to trip between stations s_1 and s_2 that begins at interval t_1 and ends at interval t_2. Each participant (rider or driver) upon registering in the system provides information on their origin and destination stations and their travel time window, which is bounded from below by the earliest departure time and from above by the latest arrival time of their trip.

For each rider r, a graph G_r can be constructed based on these parameters. A link on this graph exists if it is spatiotemporally accessible by at least one driver (i.e., both the rider and at least one driver can travel on that link). The algorithm searches on this graph for a minimum cost path that starts from the origin station of the rider and ends at his/her destination station. We define the cost of a path as a weighted linear combination of the in-vehicle travel time, waiting travel time, and number of transfers. Note that a rider can use multiple vehicles/modes of transportation to accomplish his/her trip.

The algorithm tries to find the best itinerary for rider r, by first topologically sorting the graph G_r, and then searching on this graph using a DP algorithm for an optimal path. The Bellman equation for the DP algorithm is presented in equation (6.1). $V(n_j, d)$ in this equation is the value of the minimum cost path from node 1 in the topologically ordered graph, to node n_j, with d being the last driver in the set of drivers that form the itinerary of the rider. The cost of each path is a linear combination of the travel time between nodes n_i and n_j, $C(n_i, n_j)$, which is itself a linear function of the in-vehicle and waiting travel times, and a fixed penalty, C_T, for each transfer. Set D_i^{in} denotes the set of drivers who enter node i, and set DN_j denotes a set of tuples (n_i, d), such that there is a link for driver d from node n_i to node n_j. Finally, set $ED(n_i, d')$ contains a list of drivers on the optimal path to node n_i, excluding the driver on the last link. The purpose of using this set is to prevent an itinerary to switch between the same set of drivers (e.g., traveling with driver 1, then transferring to driver 2, and then transferring to driver 1 again.) For more details on this algorithm refer to Masoud and Jayakrishnan (2015*b*).

$$V(n_j, d) = min_{n_i:(n_i,d)\in DN_j}\Big(min_{d'\in D_i^{in}\setminus ED(n_i,d')}(v(n_i, d') + C_T 1_{\{d\neq d'\}}) + C(n_i, n_j)\Big) \tag{6.1}$$

The above algorithm has multiple practical advantages in the context of P2P ride exchange. First, the algorithm can be easily modified to make it suitable for one-to-one matching, either by setting C_T to a large value, or by modifying the input sets. Second, the algorithm can be used to run a many-to-many ridesharing system, by taking into consideration the previously matched drivers when building the graph G_r for rider r. Third, the trees generated during the iterations of DP can be stored and used later to retrieve additional feasible itineraries for a rider in case he/she is a candidate to trade his/her itinerary for a sub-optimal one in P2P ride exchange.

6.4 Peer-to-Peer Ride Exchange

Dynamic ridesharing systems should have the capability of matching riders and drivers in real-time. Since participants in a dynamic ridesharing system announce their trips not long before they are ready for departure, the attempt to find a match for them should start as soon as the trip announcement is received by running the DP algorithm described in section 6.3. If all the itineraries generated by the algorithm are infeasible due to their conflicts with itineraries of the previously assigned riders (i.e., if the itineraries use the same drivers, but through different paths), then the system evaluates the possibility of a trade. In this section, we show through an example the benefits of a P2P exchange program, discuss the conditions under which trade can happen, and devise a mechanism that ensures a fair trade.

Let P^r denote the set of itineraries for rider r. Each itinerary has a value that is determined by a pre-specified objective function (the DP objective function), based on which the itineraries within P^r are ranked. Let p_i^r denote the i^{th} itinerary of rider r, and $d(p_i^r)$ denote the set of drivers who contribute to itinerary p_i^r. Note that there is no need to know all members of set P^r in advance, but we will generate them as (and if) needed. Furthermore, let p_k denote the itinerary of the assigned driver k.

Once rider r joins the system, the system uses the DP algorithm to generate a set of trees from which members of set P_r can be retrieved. The system starts by evaluating members of set P^r in order of their ranking. If an itinerary with no conflicts with the itineraries of previously matched drivers is found, this itinerary will be assigned to rider r. If the system exhausts all members of set P_r, and is not successful in finding a non-conflicting itinerary for rider r, then it considers the possibility of a trade.

Assume that rider 1 enters the system, and has two itineraries: $P_1 = \left\{p_1^1, p_2^1\right\}$, where $d(p_1^1) = \{d_1\}$ and $d(p_2^1) = \{d_2\}$. The left hand side picture in Figure 6.1 shows the rider

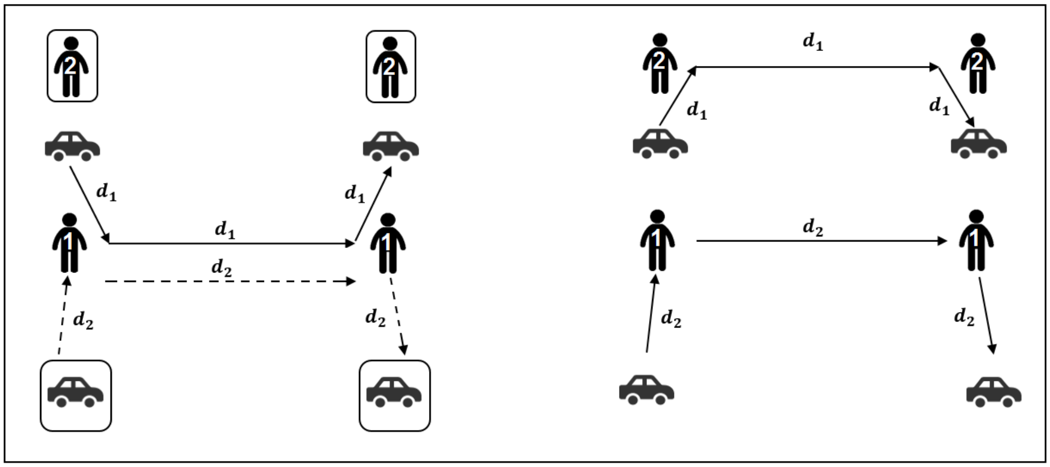


Figure 6.1: An example of a successful trade. Solid and dashed lines represent proposed and potential itineraries, respectively.

and his itinerary set. Assuming that the minimum cost itinerary for this rider is the first one, this itinerary will be announced to both rider 1 and driver 1. Next, rider 2 joins the system. Because rider 1's itinerary has been fixed, there are no feasible itineraries for rider 2. However, rider 2 has a chance to buy rider 1's itinerary if rider 1 has not started his trip yet. The right hand side picture in Figure 6.1 shows this scenario after the trade. In this trade, rider 2 buys rider 1's assigned itinerary, and by doing so liberates driver 1, who in turn forms a feasible itinerary for rider 2. Rider 1 switches to a less convenient itinerary (with driver 2) in exchange for a monetary compensation. This trade's contribution to customer retention is double-folded. Not only are both riders served, but now both drivers are participating in the system as well.

Note that it is possible to obtain the same optimal solution by solving a many-to-many ride-matching problem that is capable of considering both riders at the same time. There are, however, two issues with such an approach: (1) An optimal matching algorithm that could consider both riders at the same time is computationally harder to solve (specially for real-world size problems), and therefore might not be able to yield solutions in real-time. (2) Even if the system is equipped with a many-to-many ride-matching algorithm that can yield solutions in a reasonable period of time, the information on the two drivers and the two riders need to be available in advance for the many-to-many matching problem to generate the solution that can serve both riders.

The system studies the possibility of a trade if the following three conditions hold. First, the buyer does not have any feasible itineraries; second, the seller has an alternative feasible itinerary to his current one; and third, both parties will be better off with the trade than without it.

The monetary transfer from the buyer to the seller covers the additional cost the seller has to endure due to itinerary-switching. This cost includes the additional monetary cost due to a potentially increased travel distance, and a compensation to the seller for a potentially increased travel time. A proportion of this money will be used by the system operator to cover the cost of the seller's new itinerary that is now more expensive, and the rest will be transferred to the seller himself.

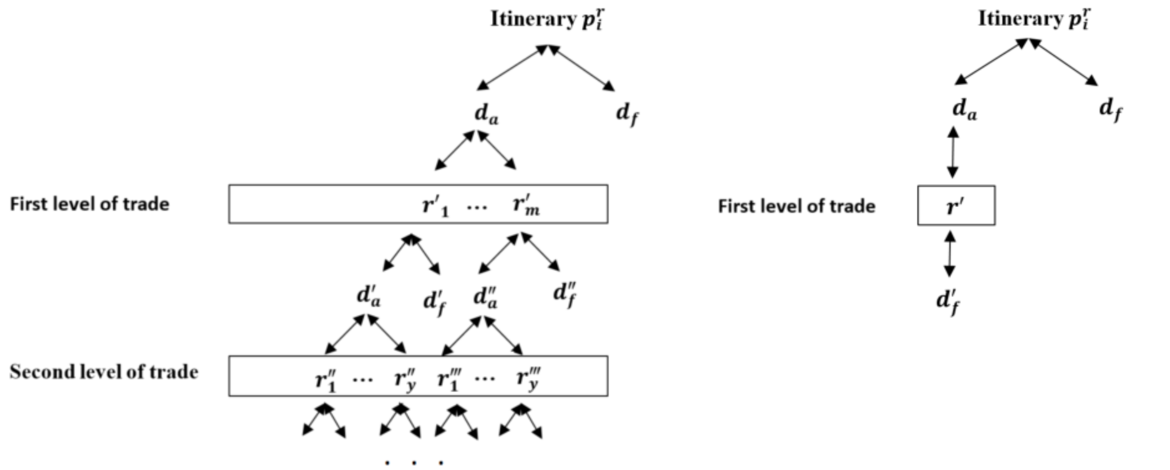
6.4.1 The Scope of the Trade

Assume a set of itineraries P^r for rider r. Drivers contributing to itinerary i are stored in set $d(p_i^r)$. Let us divide members of set $d(p_i^r)$ into two mutually exclusive sets, $d_a(p_i^r)$ and $d_f(p_i^r)$. Drivers in set d_a have been previously assigned to other riders, but their corresponding riders' trips have not started yet. Drivers in set d_f are free, and have not been assigned to any riders. The necessary condition for rider r to have a feasible itinerary is for at least one of the driver sets $d_a(p_i^r)$ and $d_f(p_i^r)$ to be non-empty. As the sufficient condition for p_i^r to be a feasible itinerary for rider r, one of the following conditions should hold: (1)$d_a(p_i^r) = \emptyset$,i.e., none of the drivers that contribute to the itinerary are assigned to other riders, and (2) $\forall k \in d_a(p_i^r), p_i^r(k) \in p_k$,i.e., drivers in set $d_a(p_i^r)$ can still follow their previously assigned itineraries. If none of these two conditions hold, then the system tries to find a good candidate for a trade between the assigned riders.

To find the candidates for a trade, the system has to first identify the itinerary that rider r is interested in. It starts from the best itinerary, i.e., p_1^r, and moves to the next itinerary if the trade on the current itinerary is not possible. In order for the system to offer itinerary p_i^r to rider r, it has to liberate all the drivers in set $d_a(p_i^r)$ from their previous assignments. Therefore, the system has to find all the riders who are using these drivers, and find alternative itineraries for them as well. These riders form the sellers in the first level of trade (Figure 6.2a). In order for the system to propose an exchange to a rider r' in the first level of trade, it should find an alternative itinerary for this rider first. This task can be accomplished by identifying the set of assigned drivers for the rider (d_a'), finding the rest of the riders whose itineraries are affected by these drivers, and finally finding alternative itineraries for them as well. This procedure continues until the system reaches a level of trade where all riders have itineraries with free drivers (or previously assigned drivers with non-conflicting assignments).

The system will then start proposing trades to riders, starting from those in the last level of trade. In order for a trade to be approved at any level, all the riders at that level should approve the trades proposed to them. For the n^{th} level of trade to take place, the trade at level $n+1$ should have been approved. Once all riders in a given level approve the proposed trades, the system can move to the next (higher) level of trade (moving upwards in Figure 6.2a). Therefore, it is clear that the more the levels of trade there are, the less likely it is for rider r to obtain itinerary p_i^r.

Another complication is that even if only one rider does not agree to the trade at a certain level, the trade cannot happen. In this case all the riders in the same and lower levels who have agreed to the trades proposed to them have to go back to their previous itineraries. Therefore, in order to simplify this procedure and make it easy to implement in practice, this chapter only considers trades in settings where the level of trade is limited to 1, and the number of riders in the first level of trade is limited to 1 as well, i.e., the set of assigned



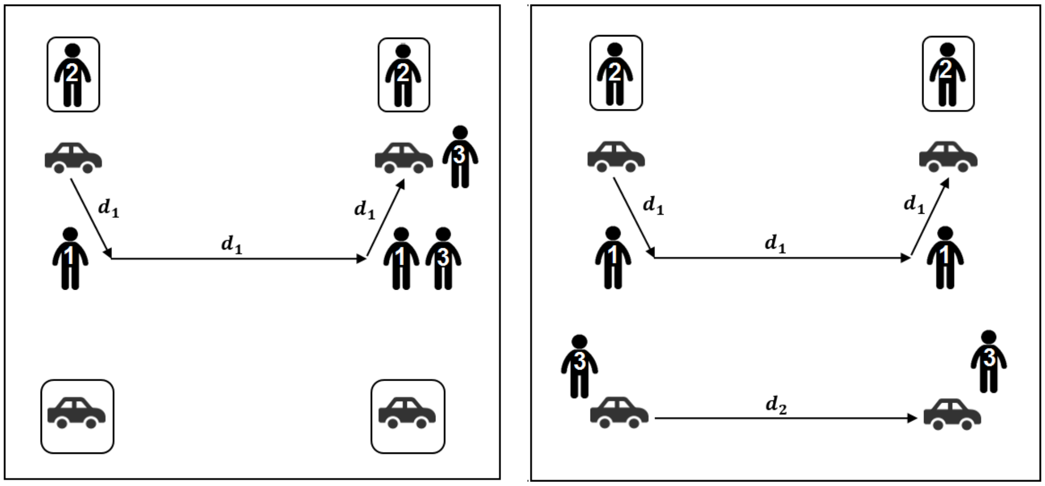
(a) General levels of trade (b) Level of trade covered in this chapter

Figure 6.2: Levels of trade

drivers affect only the itinerary of a single previously assigned rider (Figure 6.2b). These simplifications limit the trade between two individuals only: the buyer, r, and the seller, r'.

Figure 6.3 displays two examples involving multilateral trade. In the first example (Figure 6.3a), in order for the system to serve rider 2 by liberating driver 1, it must find alternative itineraries for both riders currently served by driver 1. For this to happen, rider 2 should negotiate with both riders 1 and 3, which is beyond the scope of the bilateral trade covered in this chapter. Notice that this example is still limited to the first level of trade (Figure 6.2a). Furthermore, even if a multilateral trade mechanism was available, in order for the trade to happen, both riders 1 and 3 should have agreed to the trade.

Figure 6.3b demonstrates an example of a simple scenario involving riders beyond the first level of trade. In this example, two alternative itineraries are available for rider 1, just as in Figure 6.1. When rider 2 joins the system, he is interested in purchasing the itinerary assigned to rider 1. In this example, however, driver 2 has been matched with rider 3, who belongs to the second level of trade. Therefore, for the trade to happen, the system should find an alternative itinerary for rider 3 first.



(a) More complex trade at the first level (b) Trade at the second level

Figure 6.3: Examples of multilateral trade

6.4.2 P2P Ride Exchange Mechanism

Besides ensuring that the trade makes both parties better off, the designer (operator) should also ensure that the trading parties cannot manipulate the outcome of the trade. Since both the buyer and the seller hold private information not known to the operator (e.g., their value of time (VOT)), this could lead to an inefficient outcome. This issue is addressed by modeling the trade from a mechanism design perspective. Informally, a mechanism is a method that defines rules for a game with incomplete information (Bayesian game) to influence agents' behavior and reach a particular goal, which in this case is efficiency maximization. Excellent introductions to mechanism design can be found in (Mas-Colell et al., 1995; Nisan and Ronen, 2001). The reader is invited to read Chapter 3 for a brief introduction on the mechanism design concepts employed in this section.

Let $I = 1,\ldots,n$ be the set of agents. Each agent has a type (value of time),$\theta_i \in \Theta_i$ which is private. $\Theta = \times_{(i \in I)} \Theta_i$ is the type profile set. Agent i has the (quasi-linear) utility function $u_i(\theta_i, \theta_{-i}; \theta_i) = v_i\Big(k(\theta_i, \theta_{-i}); \theta_i\Big) - p_i(\theta_i, \theta_{-i})$. Where $v_i(.)$ is his valuation and $p_i(.)$ is the price charged to him. The types before the semicolon are the types announced to the designer, while the type at its right is the agent i's actual type. A (direct revelation) mechanism is composed of two interrelated functions. The first is an allocation function $k : \Theta \to K$ that maps the type space to an outcome set K. That is, for every announced type profile, an allocation $k(\theta) \in K$ is given. In the case of a bilateral trade, K is composed of the two allocations (trading states): either there is trade or there is not. The allocation rule that maximizes the sum of agents' valuations is the efficient allocation rule, $k^*(\theta) \in K$. Secondly, there is a payment function $p : \Theta \to R^N$. This function assigns a transfer amount to every agent i in accordance with its announced type θ_i.

The trade is modeled as a bilateral trade with private information (Myerson and Satterthwaite, 1983; Hagerty and Rogerson, 1987). We follow a "robust" implementation approach in which the designer attempts to maximize the expected surplus from the trade. It is assumed that the designer has a prior on the private information from agents, but the agents themselves do not have a prior of the other agents' type, unlike in common truthfulness concepts such as Bayesian Incentive Compatibility (Myerson and Satterthwaite, 1983). The designer proceeds to find the optimal posted price that maximizes expected surplus based on that information. This framework is very convenient for our purposes, since the mechanism has to be designed far in advance, with no previous experience or learning on trading outcomes from either the users' or the designer's part, while, at the same time, it aims an increase in the number of served riders to achieve user permanence in the system. Hagerty and Rogerson (1987) has shown that in this kind of a bilateral trading setting, any DSIC mechanism is a posted-price mechanism.

Let $I = 1, 2$ be the set of agents, $i = 1$ being the seller and $i = 2$ being the buyer. Each agent i has type $v_i = [\underline{v_i}, \overline{v_i}]$. These types are drawn from an empirical VOT distribution estimated from a survey on households conducted in Stockholm, Sweden in 2005 (Abou-Zeid et al., 2010). In that research, the Stated Preferences (SP) choice scenarios are composed of car alternatives that differ on attributes such as travel times and travel costs. Since only the main statistics are available in the publication, the distribution is recalibrated as a lognormal distribution given these statistics. Its parameters are location $\mu = 2.16$ and scale $\sigma = 0.40$.

The mechanism lies in the space $(q,p) \in [0,1] \times R$, where q is the probability of trade and p is the payment from the buyer to the seller. For clarity in the exposition, we use the following change in notation $c_1 \overset{def}{=} -v_1$. c_1 is the opportunity cost of the seller. By definition, the bilateral trading setting satisfies the strict budget balance property, thus the seller has utility $u_1 = p - c_1 q$ and the buyer $u_2 = v_2 q - p$. Both agents are proposed the price p and if both agree, the trade takes place. This occurs when $v_2 > p > c_1$. The surplus of such as trade is $w\Big((q,p);\theta\Big) = (v_2 - c_1)q$.

Instead of valuing an object by a scalar as in the original bilateral trading environment, riders value their allocation (assigned ride) by its generalized cost, which is an affine transformation of their private type. For a rider i, this cost is the product of the travel time t_{ri} and the sum of the value of time θ_i, plus the fare per unit of time c_{ri}. These valuations are normalized with regard to the initial situation (no trade) to fit the bilateral trading original setting: c_1 and v_2 are in fact the valuation difference between states "trade" and "no trade". When there is a trade, $c_1(\theta_1)= \theta_1(t'_{r1} - t_{r1}) + c'_{r1}t'_{r1} - c_{r1}t_{r1}$ and $v_2(\theta_2) = \theta_2(t_{out} - t_{r2}) + c_{out}t_{out} - c_{r2}t_{r2}$. Here, t_{r1}, t'_{r1}, t_{r2}, and t_{out} refer to the travel time of the seller's current and new itineraries, travel time of the buyer's itinerary, and the travel time of the buyer's outside alternative, respectively. c_{r1}, c'_{r1}, c_{r2} and c_{out} are the costs per unit time of seller initial ride, seller alternative proposed ride, buyer proposed ride and the outside option cost to the buyer.

These time and cost variables have bounds and relative magnitudes. $t_{out} \leq t_{r2}$ since we consider the outside option to use the shortest path between buyer's origin and destination. We assume that cost of the rideshare option to the buyer is less than that of the outside option, i.e., $c_{out}t_{out} - c_{r2}t_{r2} \geq 0$; Otherwise the buyer would not have selected to use the rideshare option. This assumption leads $v_2 = [\underline{v_2}, \overline{v_2}]$ to be positive in our analysis. Note that as θ_2 increases, v_2 decreases, so $\underline{v_2} = max\Big(0, v_2(\underline{\theta_2})\Big)$ and $\overline{v_2} = v_2(\overline{\theta_2})$. Since the seller is offered a longer ride than the one he holds, $t'_{r1} \geq t_{r1}$ and $c'_{r1}t'_{r1} - c_{r1}t_{r1} \geq 0$, with a high probability. If due to higher number of riders involved in the new itinerary c'_{r1} becomes smaller than c_{r1}, the system will set $c'_{r1}t'_{r1} - c_{r1}t_{r1} = 0$. In this way, $c_1 \in [\underline{c_1}, \overline{c_1}] \geq 0$ and it is increasing with θ_1. Its bounds are $\underline{c_1} = c_1(\underline{\theta_1})$ and $\overline{c_1} = c_1(\overline{\theta_1})$. Without loss of generality, we assume $\underline{\theta_1} = \underline{\theta_2} = \underline{\theta}$ and $\overline{\theta_1} = \overline{\theta_2} = \overline{\theta}$. Trade is only possible when $\underline{v_2} \geq \overline{c_2}$. The next proposition formally presents the bilateral trade mechanism for a given price p and its properties:

Proposition 6.1. *The posted-price mechanism with price p for the bilateral trade setting is a revelation mechanism (q,p) such that:*

$$(q,p) = \begin{cases} (1,p) & v_2 > p > c_1 \\ (0,0) & otherwise \end{cases}$$

This mechanism is DSIC, EPIR, SBB and guarantees the following expected surplus $W(p)$:

$$W(p) = \iint_{(c_1,v_2)=\left(c_1(\underline{\theta_1}),p\right)}^{(c_1,v_2)=\left(p,v_2(\underline{\theta_2})\right)} (v_2 - c_1)\phi(c_1,v_2)dv_2dc_1 \tag{6.2}$$

Proof. Let the strategy of the seller s_{seller} be: sell if $p > c_1$ and do not sell if $p \leq c_1$. This is the only DSIC strategy for the seller (we assume that the seller prefers to keep the object when the gain is zero). Suppose that the seller follows a different strategy s'_{seller}: sell only when $p - c_1 > \epsilon, \epsilon > 0$. Then the seller would lose the opportunity to make profit $p - c_1$ when $c_1 + \epsilon > p > c_1$. Equivalently, when ϵ is negative, the seller would incur a loss of $c_1 - p$ when $c_1 + \epsilon < p < c_1$.

Moreover, s_{seller} is the only strategy that is EPIR. Seller's payoff from not participating in the mechanism is zero. Following the same reasoning as above on incentive compatibility, s_{seller} is the only strategy that provides a non-negative profit for every buyer and seller type. The same reasoning applies for the buyer, with strategy s_{buyer}: buy if $v_2 > p$ and do not buy if $v_2 \leq p$.

The mechanism can be easily seen to be strictly budget-balanced (SBB). When there is no trade, the monetary transfer is zero. When there is a trade, the positive price paid by the buyer goes to the seller. There is no waste in the numeraire in either case, which ensures the SBB property.

The expected welfare $W(p)$ is the integral of welfare function weighted by the joint type probability distribution function. The bounds determine the entire valuation range over which the trade happens and therefore captures the cases when there is positive surplus from a trade. When a trade does not happen, the surplus is zero. Since agents' VOT distributions are assumed to be independent, the VOT joint distribution ϕ is the product of the two marginal distributions. □

The problem the designer faces is to set an optimal price p^* which maximizes the probability of trade, and thus maximizes the total surplus. In the present study, the optimal price p^* is found by maximizing $W(p)$ with linear search in price intervals of 5 cents.

6.4.3 Pricing

There are many factors that should be taken into account in determining the fare for ridesharing services. Setting the right fare is essential to the success of a ridesharing system, and deserves design of a separate mechanism which ensures that no incentive exists for drivers and riders to falsely report their preferences in order to affect the amount of transaction.

The fare a rider is charged in our system is made of two components. The first component is a variable, distance/time dependent fee. Assume that rider r's itinerary involves traveling on link set L^*, and that at the time of matching the rider, on each link $\ell \in L^*$, n_ℓ number of individuals (including the driver) share the same vehicle with the rider. We assume that the cost of travel on each link is equally shared by the individuals who travel on the link. Therefore, the variable fee of rider r will be $\sum_{\ell \in L^*} \frac{d_\ell}{n_\ell}$. In this equation, d_ℓ is the general cost of traveling on link ℓ, and $\sum_{\ell \in L^*} \frac{d_\ell}{n_\ell}$ is the total share of the rider from the cost. Note that the general cost of a link can be time-based, distance-based or a combination of both.

In addition to a variable component, the fare also has a fixed component. Since drivers may have to divert from their shortest/preferred paths in order to accommodate riders, they need to be compensated for the extra travel. We calculate the base fare based on the average extra travel time drivers have to spend in the network, assuming an average speed of 40 mph, and a payment of 60 cents per mile. These fares could vary for different times the day, and days of the week, based on the composition of the ridesharing system, i.e., number of participants, the driver to rider ratio, and the degree of flexibility of riders and drivers. Although a pricing scheme that can distribute fares among drivers based on their contribution to the system may be fairer, in the interest of simplicity we use the more preliminary pricing scheme introduced in this section.

6.5 Numerical Study

In order to study the impact of the P2P ride exchange mechanism on the performance of a ridesharing system, we generate and solve multiple random instances of the ridesharing problem. All results reported here are averaged over 30 runs for each problem instance.

In each problem instance, we generate a number of participants with varying ratio of riders. The origins and destinations of participants are selected based on a uniform random distribution from a pre-specified set of stations in a grid network. The earliest departure time of each participant is selected uniformly by a random distribution within a certain "departure period". The travel time budget of each participant is determined as a factor of his/her shortest path travel time. The latest arrival time of a participants at his/her destination station is then computed as the sum of the participant's earliest departure time and travel time budget. All these parameters impact the level of spatiotemporal proximity between trips.

For each participant, a VOT is drawn from the lognormal distribution described earlier. Each individual is assumed to have a separate transportation alternative outside of the system, with a travel time equal to the shortest path travel time between the individual's origin and destination. The unit distance-based cost of the outside alternative is assumed to be equal to that of the ridesharing system.

We solve each problem instance using three different ridesharing implementation strategies. The first strategy referred to as "one-to-one", matches a single rider with a single driver. The second strategy referred to as "one-to-many" allows a driver to carry multiple riders. Riders complete their trips in a single vehicle. The last implementation strategy referred to as "many-to-many" allows each driver to carry multiple riders, and each rider to transfer between multiple drivers. Note that the P2P exchange mechanism is optimal only for the one-to-one matching method. The number of additional riders served due to exchange in one-to-many and'many-to-many systems is only a lower bound and may increase if we use a more sophisticated mechanism that can include higher levels of trade.

In this section, we perform sensitivity analysis over system parameters, namely the number of participants, ratio of riders, travel time budget factor, and number of stations. Through this analysis, we study the impact of different parameter values on the percentage increase in the number of matched riders (to which we refer as the *exchange rate*). Finally, we generate different ridesharing scenarios with different levels of spatiotemporal proximity between trips, and use statistical tests to confirm whether the observed difference in the exchange rates in these scenarios is statistically significant.

6.5.1 Base Fares

In order for a rider to decide whether to participate in a trade or not, he/she should have information on the cost of the proposed itinerary. As discussed in section 6.4.3, the cost of an itinerary entails a variable, route-dependent cost, and a fixed cost. In this section, we demonstrate for certain parameter values how this fixed cost is calculated, and how it may differ from hour to hour or day to day.

Figure 6.4 displays the base fares charged to riders and payed to drivers in a ridesharing system at different levels of trip spatiotemporal proximity. As mentioned in the previous section, this study uses the same base fares for all drivers and a different but equal base fare for all riders in a given time period, for example, during weekday morning peak hours. These fares may vary from location to location, and depend on the number of participants and system composition (number of participants, and ratio of riders to total number of participants). In this section, we show sample base fares for a system with 200 participants with departure period of 60 minutes and travel time budget factor of 1.5, under different ratios of riders and number of stations.

As figure 6.4 suggests, for a one-to-one system, the fare paid by riders is similar to the fare received by drivers, since the number of served riders and matched drivers in a one-to-one system are equal. In a one-to-many system, in which each driver carries multiple riders, the fixed fare paid by riders decreases as the ratio of served riders to matched participants increases, since now multiple riders are being served by a single driver and hence they each pay a portion of the fixed fare the driver receives. Interestingly enough, in a many-to-many system, when the ratio of riders in the problem is small, each rider pays more than what each driver receives, suggesting that riders are receiving multi-hop itineraries (i.e., transferring between drivers), especially when the number of stations is high and therefore the spatial proximity between trips is too low. After a certain point, however, the fare received by each driver surpasses the fare paid by each rider, suggesting a high level of sharing.

6.5.2 Number of Participants

In this section, we study the performance of the P2P ride exchange mechanism under different participation rates, and different numbers of stations. The problem instances have been generated in grid networks of different sizes (9, 25 and 49 stations), departure period of one hour, and 200, 300, and 500 number of participants. For a given number of participants, number of riders and drivers are assumed to be equal, since a rider to driver ratio of close to 1 is where a ridesharing system yields the highest matching rate (Masoud and Jayakrishnan, 2015*a*). Figure 6.5 shows the initial matching rate of riders, and the exchange rate under different implementations of the system.

The results suggest that under all implementation strategies, both the initial matching rate and the exchange rate are positively correlated with the participation rate, i.e., the higher the number of participants, the higher the performance of the system and the exchange rate.

Another general observation is that with all matching methods, lower numbers of stations result in higher initial matching rates, and higher exchange rates. This result is intuitive, since participants have to choose their origins and destinations from the set of stations. Therefore, a lower number of stations results in more participants sharing the same ori-

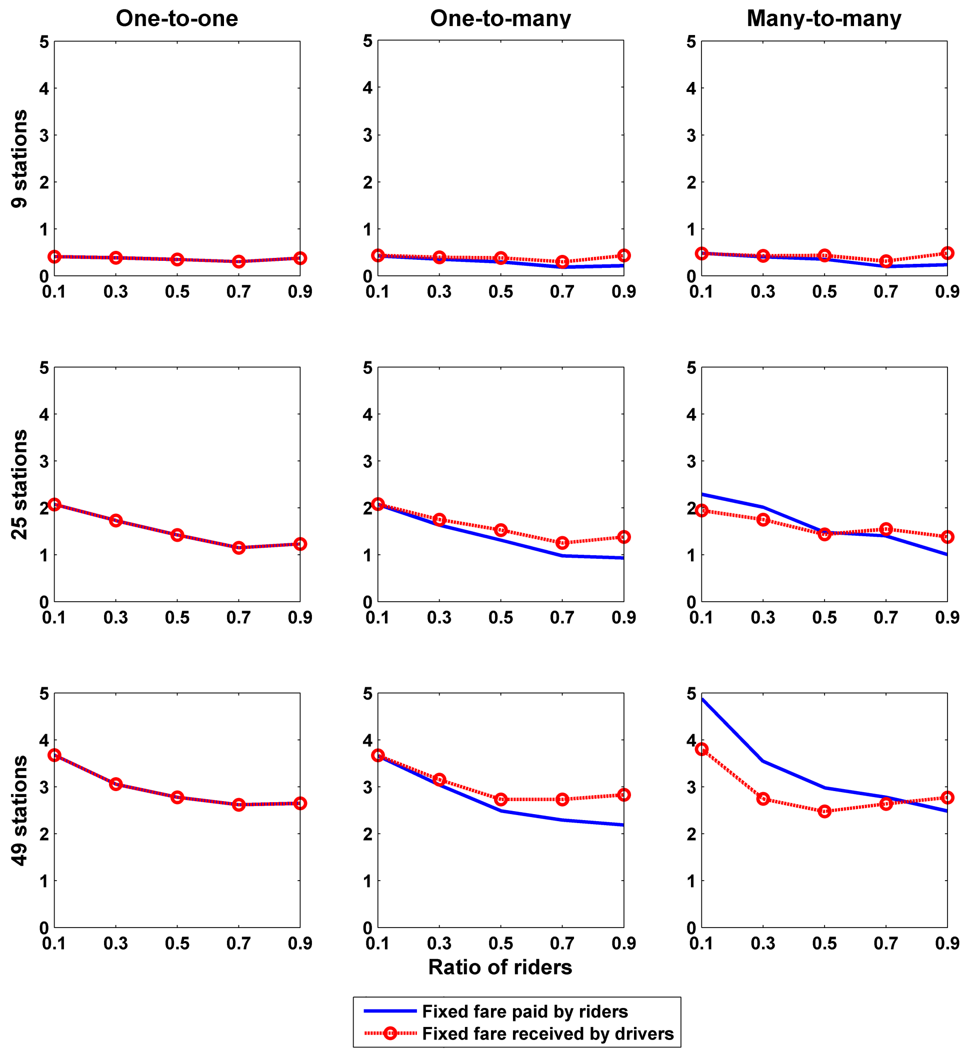


Figure 6.4: Base fares for ridesharing systems with different levels of spatiotemporal proximity between trips.

gin and/or destination stations, and therefore a higher spatial proximity between trips. Higher spatial proximity also suggests a higher probability of finding a driver that can serve the seller in the ride exchange, and therefore higher success rate in the exchange.

Despite this general trend, the impact of spatial proximity depends on the number of participants as well. For a one-to-one system with 300 participants, the exchange rate increases as we increase the spatial proximity by moving from from 49 to 25, and finally 9 stations. With 500 participants, however, it takes lower levels of spatial proximity than 49 stations to impact the exchange rate since a higher number of participants by itself increases the proximity of trips, to some degree.

The one-to-one matching method is the only method for which the result of the exchange mechanism is optimal; therefore, the matching rate for such a system is the highest. Another reason why the P2P exchange scheme may serve one-to-one ridesharing systems well is that in such systems each driver serves a single rider (if matched), which increases the likelihood of liberating the desired driver, and requires negotiations only with a single seller.

In one-to-many and many-to-many systems there are trades that expand to higher levels than the first level of trade, and therefore are not successfully completed. This is one reason behind the lower exchange rates for these systems, compared to the one-to-one system. Another reason for lower exchange rates in one-to-many and many-to-many systems is that the exchange rate measures the percentage increase in the matching rate. Even if the same additional number of riders are served due to exchanges in all implementation strategies, the many-to-many and one-to-many systems will show lower exchange rates because their initial matching rates are higher.

Figure 6.6 shows the driver matching rates and the impact of the exchange mechanism on the increase in the percentage of matched drivers. In general, this figure follows similar

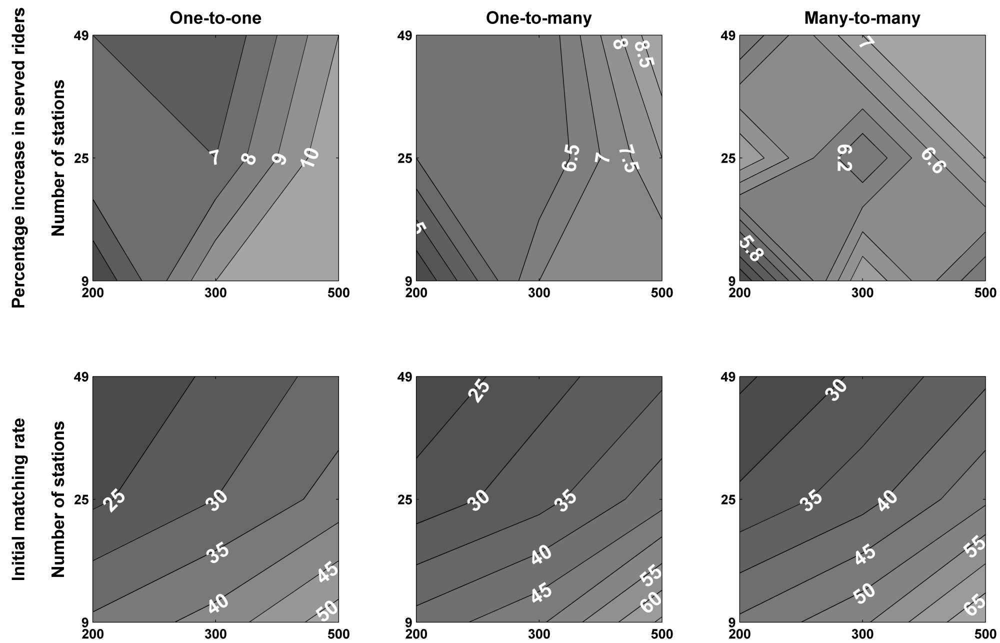


Figure 6.5: Initial rider matching and exchange rates under different number of stations and participants.

trends to Figure 6.5. In a many-to-many system the driver exchange rate is higher than the rider exchange rate, which implies a high level of sharing before the exchange.

6.5.3 System Composition

The problem instances in this section include 500 number of participants generated uniformly during a departure period of one hour, in grid networks with 25 stations. Number of riders are changed from 50 to 450 (and drivers from 450 to 50) in 100 increments, in order to study the impact of system composition on the performance of the exchange mechanism.

Figure 6.7 summarizes the results. For all ridesharing implementations, as the ratio of riders to participants increases (and ratio of drivers to riders decreases), the rider matching rate experiences a declining trend. At rider ratio of 0.1 (50 riders and 450 drivers),

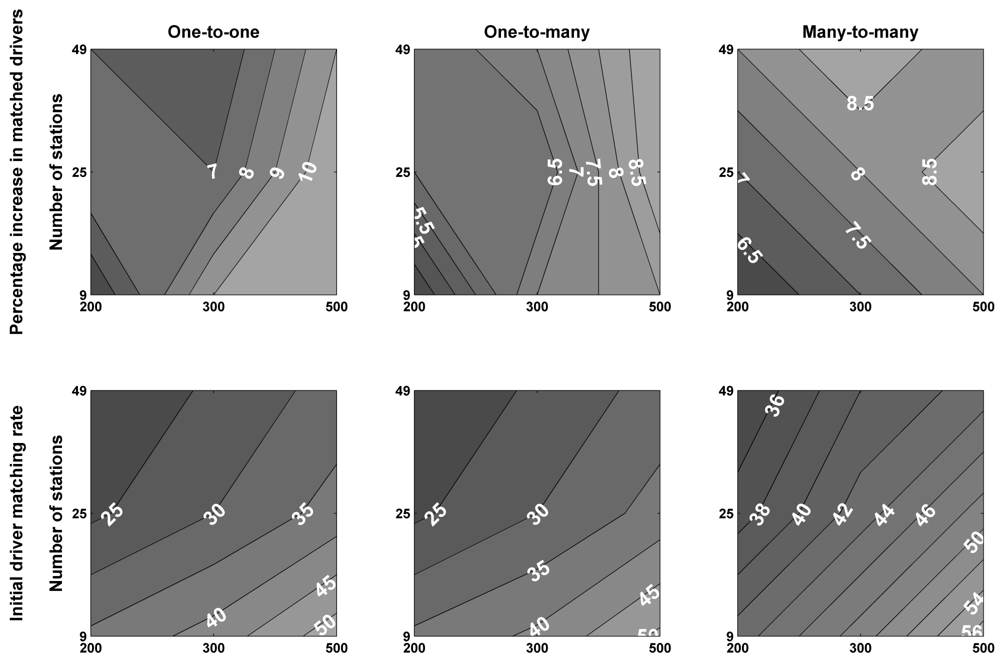


Figure 6.6: Initial driver matching and exchange rates under different number of stations and participants.

most of the 50 riders can be matched, due to the abundance of supply (i.e., drivers). As we move along the horizontal axes, the increase in demand is met with a decrease in supply, which leads to a decreasing trend in the rider matching rate. At rider ratio of 0.9 (450 riders and 50 drivers) the matching rate of riders is the lowest, due to the lack of sufficient supply.

The trend in the driver matching rate is the opposite of the trend in the rider matching rate. At lower rider ratios where there are a few riders and many more drivers, the driver matching rate is low, since a low percentage of drivers is enough to satisfy the demand. As the rider ratio increases, the demand becomes higher than supply, and so the driver utilization rate increases.

Another general trend among matching rates is the bell-shaped form of the rider and driver exchange rates. In the beginning, when rider to driver ratio is low, a high percentage of demand is satisfied, and there is not much need for exchange, hence the low exchange rate. At high ratios of riders to system participants, the rider matching rate is small, but the driver matching rate is high, and therefore there are not many free drivers left to form alternative itineraries for the sellers. In case of a many-to-many system, an exchange mechanism that can extend to higher levels of trade could help to pick up the declining exchange rate. In the middle ranges, at rider ratio of 0.5 for the one-to-one and one-to-many systems and 0.3 for the many-to-many system, the exchange rate becomes the highest. In this range, the rider matching rate is not too high to eliminate the need for exchange, and the driver matching rate is not too high to decrease the chance of finding alternative itineraries for the sellers.

6.5.4 Departure Period

Similar to the previous section, problem instances in this section consist of 500 participants, with equal number of drivers and riders. The participants are assumed to have a travel time budget factor of 1.5, and their origins and destinations are randomly selected from 25 pre-specified stations. We increase the departure periods of individuals from 15 minutes to 60 minutes, in order to study the impact of higher temporal proximity among trips on the performance of the exchange mechanism.

Figure 6.8 shows that under all ridesharing implementations, both the initial rider and driver matching and exchange rates increase with the temporal proximity among trips (i.e., as the departure period becomes smaller.) This is not surprising, since higher temporal proximity among trips leads to higher probability of finding a match in the first place, and a higher probability of finding alternative itineraries for any seller, in case an exchange is required.

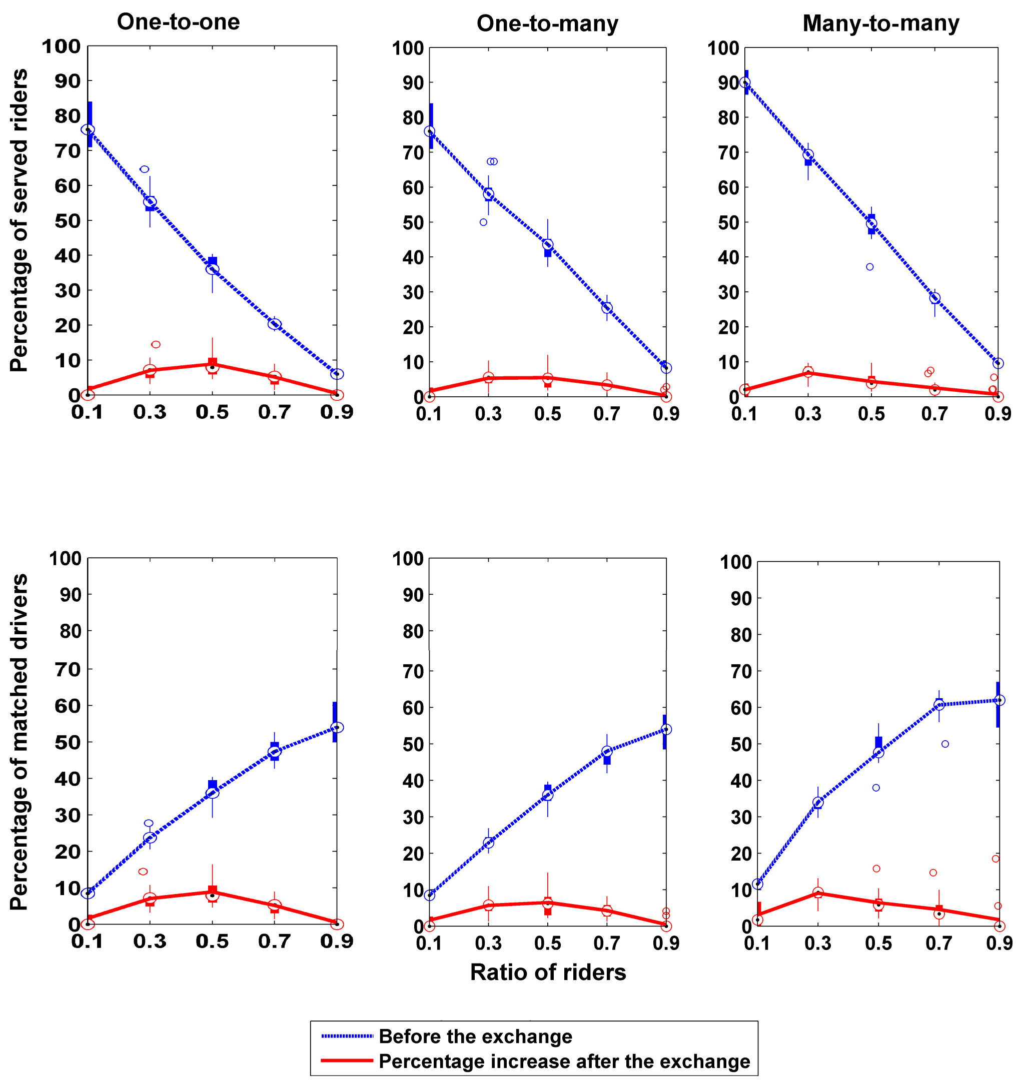


Figure 6.7: Matching and exchange rates under different ratio of riders.

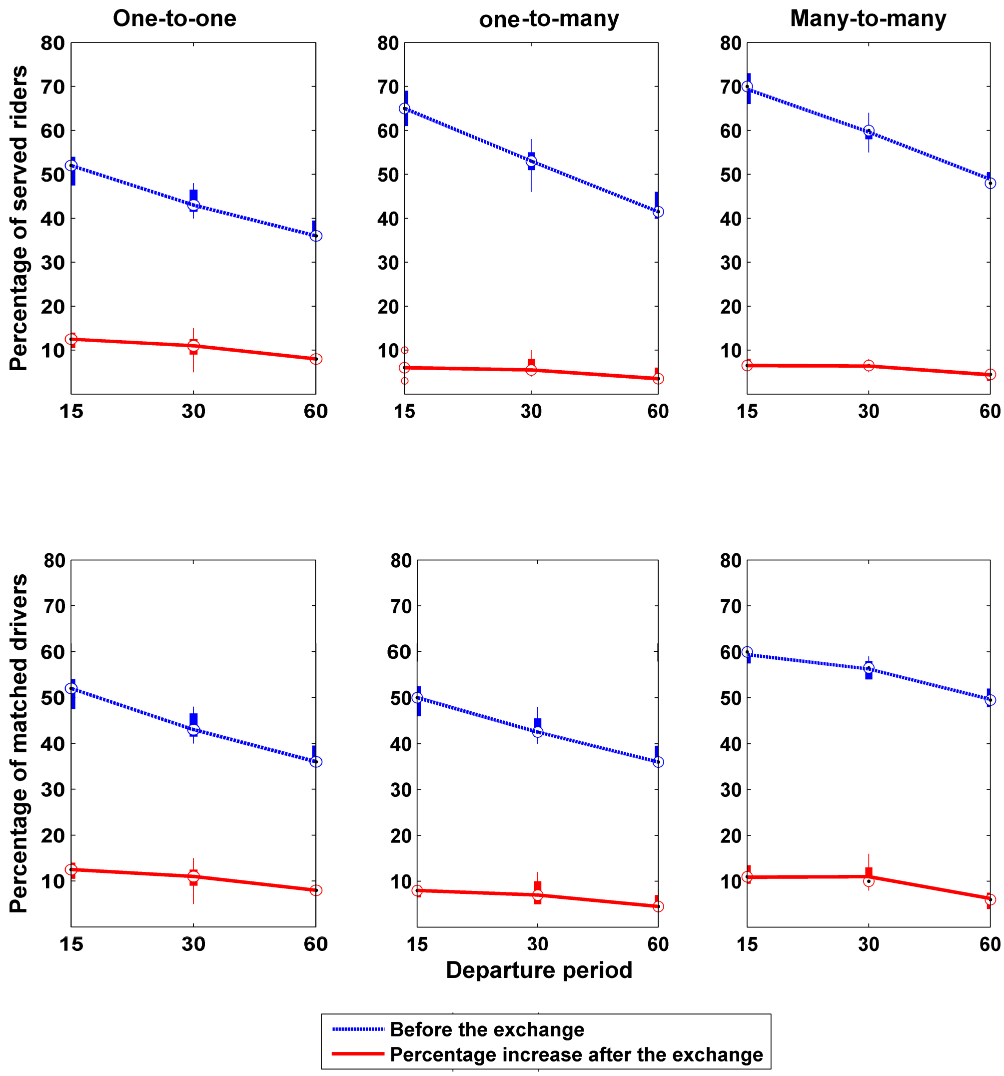


Figure 6.8: Percentage of riders and drivers matched before and by P2P ride exchange under different departure periods.

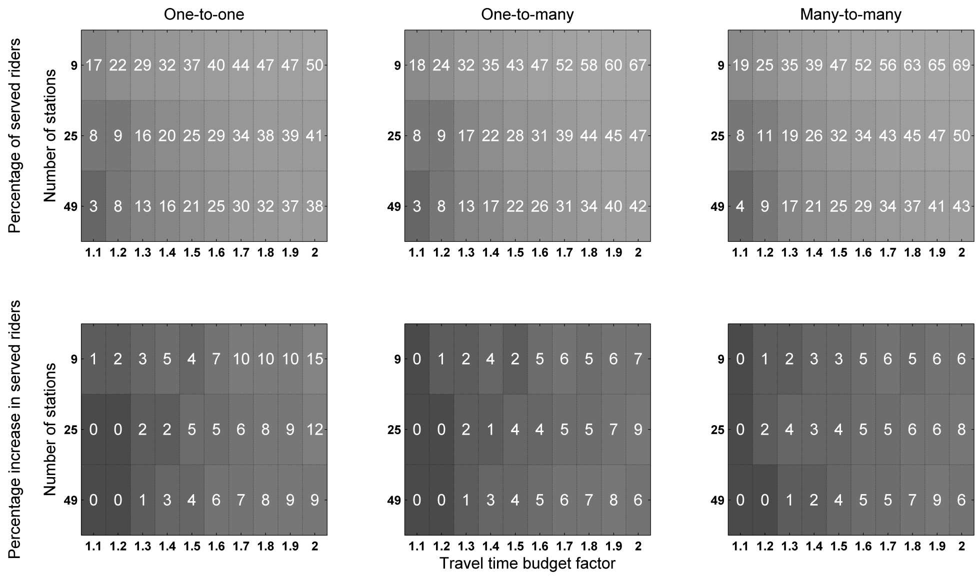


Figure 6.9: Rider matching and exchange rates as a function of travel time budget factor of participants.

6.5.5 Travel Time Budget Factor

Experiments in this section include 200 participants, with equal number of riders and drivers, and a departure period of 60 minutes. Figure 6.9 shows the initial rider matching rate, and the percentage increase in the number of served riders under different travel time budget factors for participants. This figure suggests that in general the matching rate as well as the exchange rate increase with the travel time budget factor, regardless of the change in the proximity of trips. Figure 6.10 suggests the same results for drivers.

6.5.6 Statistical Analysis

Table 6.1 lists 10 ridesharing scenarios sorted in an increasing order of exchange rate. Scenarios have different levels of spatiotemporal proximity among trips, and are all generated for a one-to-one matching method, since it is the only matching method for which the exchange mechanism is optimal.

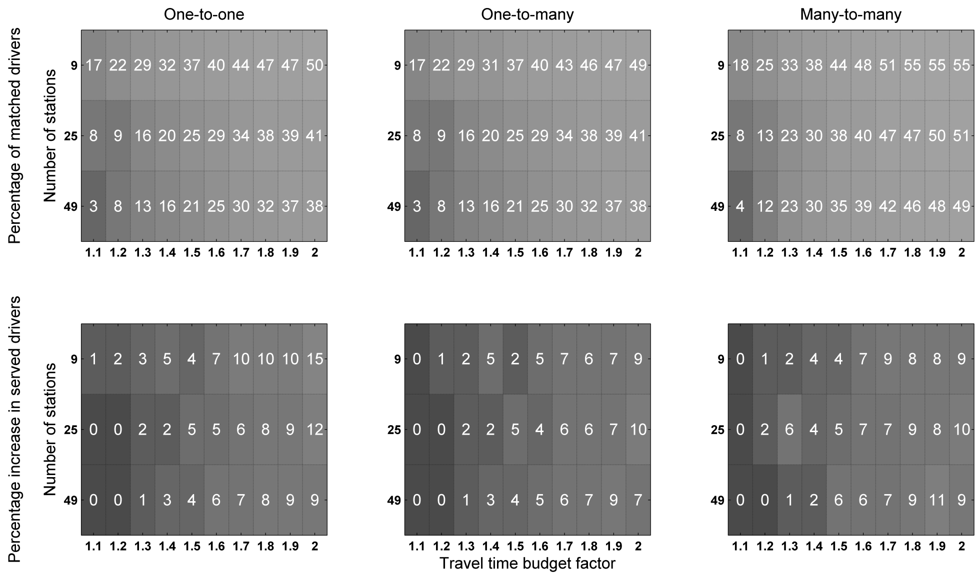


Figure 6.10: Driver matching an exchange rates as a function of travel time budget factor of participants.

As demonstrated in the previous sections, when the spatiotemporal proximity of trips is too low or too high, the exchange rate is small. When the spatiotemporal proximity among trips is too low, the initial matching rate is too low, and so is the exchange rate due to lack of supply. When the spatiotemporal proximity is too high, a high percentage of demand is served, and there is not much room left for improvement by making exchanges. In the middle ranges, where the spatiotemporal proximity among trips is moderate, is where the exchange rate is the highest. Table 6.1 lists scenarios that cover the entire range. The mean and standard deviation of the matching and exchange rates for each scenario are presented in this table. For each scenario, we have also noted the total social surplus of the system obtained due to exchange, averaged over the 30 generated instances for each scenario.

Figure 6.11 demonstrates the statistical significance of the difference in exchange rates among scenarios. For each pair of scenarios, we have used a two-sample t-test with the null hypothesis that the two scenarios come from independent random samples drawn

Scenario properties	Per. of served riders (mean,st.dev.)	Per. increase in served riders (mean,st.dev.)	Per. of retained riders	Average social surplus
(500,0.1,15,9,1.5)	(99,1.5)	(0.2,0.4)	0.6	11.33
(200,0.25,15,9,1.5)	(9,0.6)	(4.5,3.2)	11.5	10.92
(500,0.9,15,9,1.5)	(10,0.4)	(7.2,3.6)	18	8.82
(200,0.75,15,9,1.5)	(8,0.7)	(8,4.9)	20	8.66
(200,0.5,15,9,1.5)	(13,1.2)	(10,3.7)	26	9.69
(200,0.6,15,9,1.5)	(12,0.9)	(10,4)	25	9.47
(500,0.7,30,9,1.5)	(25,1.3)	(10,3)	28	9.01
(500,0.7,15,25,1.5)	(21,1.3)	(11,3.2)	28	14.34
(500,0.7,15,9,1.5)	(33,1)	(15,3)	38	9.97
(1000,0.5,30,25,1.5)	(60,1.6)	(15,1.2)	32	15.57

Table 6.1: Ridesharing instances. The scenario properties include (no. of participants, ratio of riders, departure period, no. of stations, travel time budget factor)

from two normal distributions with equal means and equal but unknown variances. The alternative hypothesis is that the two scenarios come from populations with unequal means. The null hypothesis is rejected under a 5% significance level.

Figure 6.11 shows the p-values for the two-sample t-tests for each pair of scenarios. This figure suggests that under the 5% significance level each scenario is not statistically different from one or two scenarios with higher exchange rates, but as the difference in the exchange rates among scenarios increases, the null hypothesis is rejected, and the scenarios are shown to be statistically different. This figure suggests that the difference in the exchange rates observed for different ridesharing systems with different parameter values and levels of spatiotemporal proximity among trips is statistically significant.

6.5.7 Customer Retention

The higher number of served riders due to exchange does not have a one-to-one impact on the performance of the system. It is true that the number of served riders increases only by the number of successful exchanges, but the impact on customer retention and

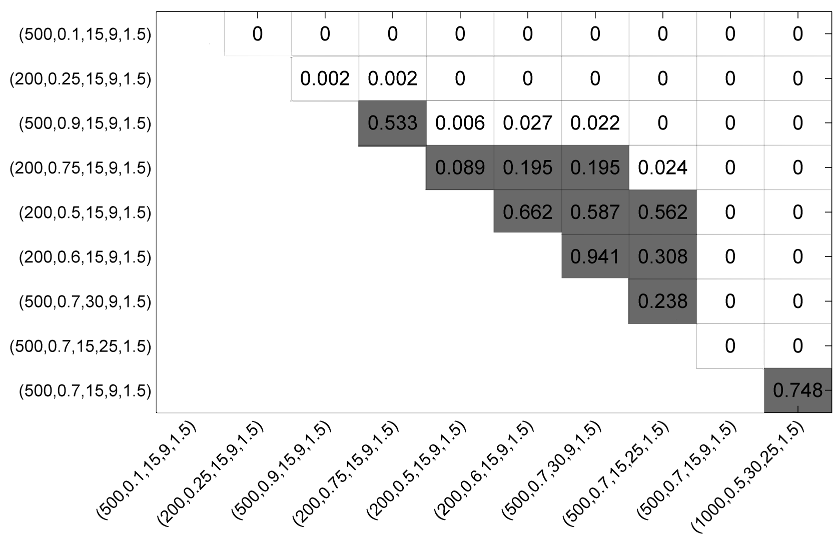


Figure 6.11: p-values between the exchange rates is scenarios listed in table 6.1

the reputation of the system should also be taken into consideration.

To study customer retention, we assume that a rider does not return to the system if he/she has three failed experiences. By simulating a three day experience for each rider, it is possible to compute the number of retained customers in a 3-day period. To conduct this analysis, we use the same set of riders in each instance of each of the scenarios listed in Table 6.1 (i.e., we change only the driver set for each problem instance in each scenario). To compute customer retention, it is assumed that a rider will consider using the system again if he/she has at least one successful experience. The percentage of retained riders due to exchange are reported in Table 6.1.

In addition to creating a positive experience for these riders and increasing the probability of them returning to the system, using P2P exchange could eliminate the possible negative WOM that could have been generated by these riders had they not been served, and could even replace them with a positive WOM.

6.5.8 Higher Levels of Trade

This study concentrated on the simplest possible scenario for trade, where there is a single buyer and a single seller. We study the impact of higher levels of trade on the exchange rate by considering a many-to-many ridesharing system for scenarios in Table 6.1. The results show an average of 20% increase in the number of served riders, over all scenarios. Numbers in table 6.1 are generated solely based on the operational feasibility of exchange (e.g., making sure for a rider who sells his/her current itinerary, there are alternative non-conflicting itineraries available), and not the monitory transactions. In addition, note that although theoretically a multi-level trade can increase the percentage of served riders significantly, moving to lower levels of trade can make the trade harder to manage and less probable to succeed, due to the requirement for all individuals involved in the trade to agree to it.

6.6 Conclusion

In this chapter, we introduced the P2P ride exchange mechanism to increase the performance level of a dynamic P2P ridesharing systems. Extensive numerical experiments suggest that this mechanism results in higher performance levels and customer retention rates than standard FCFS allocation. We demonstrated through experiments that the percentage of increase in the matching rate using P2P ride exchange is positively correlated with the number of participants, travel time budget of riders and drivers, and spatial and temporal proximity of trips. Furthermore, we showed that the exchange mechanism is more effective in terms of increasing the percentage of served riders when the ratio of number of riders to system participants is within a moderate range (not too low or too high).

We showed that the mechanism employed is robust towards selfish manipulation from users and helps in increasing ridership with minimal information requirements from users. Despite these low information requirements, an increase in the probability of getting a ride is achieved, leading to further user permanence in the system, even in situations where the system operation has not yet started and no data is available. Once more data is available, a new mechanism that takes into consideration observed user interaction may be designed.

Chapter 7

Market design: P2P Ridesharing as an alternative mode for daily commuting

7.1 Introduction

Traffic congestion in urban areas is a major concern that leads to losses of about \$260 billion nationwide or \$960 per commuter every year according to Schrank et al. (2015). Several alternatives have been put in place during the last decades to ameliorate congestion. Transportation demand management, congestion pricing, public transportation, and additional infrastructure investments such as High Occupancy Vehicle lanes (HOV) have served to reduce this externality. However, during the last decade, a historical alternative, ridesharing, has seen its revival due to the advent of smartphones and mobile internet.

Peer-to-peer (P2P) ridesharing systems suffer from a "chicken-and-egg" problem that hampers their implementation, namely of not reaching a potential critical mass that makes them a reliable alternative. One of the reasons why riders might be skeptical of leaving their vehicles at home and participanting in a ridesharing system is the uncertainty about finding a ride back home. To address this concern, we model the operations of a single-hop ridesharing system (one in which each matched driver makes a detour to carry a single passenger) with a guaranteed ride-back in the evening, targeted for commuter trips. This study is the first to model this customer attraction policy in a P2P rideshare matching and pricing scheme.

Ridesharing reduces travel cost per traveler due to sharing trip monetary costs among multiple people. On the other hand, under certain traffic conditions, HOV lanes can offer travel time savings to any vehicle with an occupancy higher than one or two passengers. Current studies on pricing for ridesharing systems have overlooked this possibility, despite its being known that carpoolers positively value travel time savings (Li et al., 2008). We thus understand P2P ridesharing as a differentiated mode that not only offers distance cost savings but also travel time savings in the presence of HOV lanes, focusing on peak hour commuting. However, potential system users must be knowledgeable of the resulting travel time savings so that they can take this factor into consideration when it comes to making mode choice decisions.

The objective of this chapter is to present a matching and pricing mechanism for a P2P ridesharing system which ensures a ride-back for the matched riders and accounts for potential HOV travel time savings. A parametric study investigates whether such a system can serve as an effective and financially self-sustainable alternative mode to commuting solo. Since there is value heterogeneity across the commuter population, the system operator must ensure that these values are elicited truthfully from users and prevent selfish manipulation of the system. We address these issues by using mechanism design solutions.

The mechanism is based on a modified version of the Vickrey-Clarke-Groves (VCG) mechanism, which maximizes social welfare and ensures truthfulness and voluntary participation from users. This chapter models the operation of a single-hop ridesharing system with a guaranteed ride back, which can be formulated as a min-cost max-flow formulation that is of polynomial nature and fast to solve using standard linear programming methods.

Unfortunately, there is a tradeoff between economic efficiency, truthfulness, budget balancedness and individual rationality. It is known (Green and Laffont, 1979) that generally there does not exist a mechanism that is guaranteed to satisfy the four properties at the same time. Actually, in practice, ridesharing systems usually depend on external subsidies to cover operational costs and make the system attractive for users. In this study, we use reserve prices to address this issue.

To assess realistic quantitative benefits of the proposed system, we present a real-world study of the Los Angeles metropolitan area. Origin-Destination demand volume data for this study comes from the Southern California Association of Governments (SCAG) model, OD travel times are queried through Google Maps API, and individual valuations of time savings and distance savings are drawn from empirical distributions.

This chapter makes several new contributions. We introduce the ride-back, enabling ridesharing to be an alternative mode to commuting. We are the first to consider HOV lane travel time savings on a P2P ridesharing (truthful) pricing scheme, which at the same time helps to control budget deficit problem. Finally, we provide a novel min-cost max-flow modeling of the problem which allows it to be solved rapidly in polynomial time which will be critical for an eventual implementation of dynamic ridesharing.

The chapter, as published in Lloret-Batlle et al. (2017), is divided as follows. Section 2 presents the review on existing literature, section 3 provides the method preliminaries, section 4 describes the mechanism, section 5 describes the data used and its treatment, section 6 contains the numerical results and finally, we state conclusions and further research.

7.2 Literature review

There has been a recent interest in using mechanism design for pricing in ridesharing operations (Kleiner et al., 2011), (Shen et al., 2013), (Kamar and Horvitz, 2009), (Zhao et al., 2014) and (Masoud and Lloret-Batlle, 2016). Most of the literature is concerned with using pricing mechanisms not only to address incentive issues but also with how to increase the probability of matching and therefore the overall efficiency of the system. These problems are common in either day-to-day ridesharing or real-time dynamic ridesharing. Alternatively, Herbawi and Weber (2011*c*), Agatz et al. (2010), Ghoseiri (2013), Masoud and Jayakrishnan (2017), and Regue et al. (2016) are some examples of studies that focus on system efficiency and performance, and leave out the pricing component.

Kleiner et al. (2011) proposes a parallel auction scheme based on a sealed-bid second price auction. They examine how the use of auctions to allocate rides can lead to more efficient outcomes than allocating rides based on Vehicle Kilometer Travelled. Shen et al. (2013) also proposes an online posted-price mechanism for centrally-managed autonomous mobility-on-demand transportations systems. Their solution is incentive compatible, individually rational and budget balanced and perfoms better than their online auction benchmark.

Kamar and Horvitz (2009) studies the viability of ridesharing plans using GPS data. The method used is a VCG-based mechanism and they address the budget deficit problem by imposing a positive budget constraint in line with Parkes et al. (2002). Agents' utility function is the sum of travel time and delayed waiting time, but they do not include distance savings. Their approach is not incentive-compatible.

Zhao et al. (2014) examines several theoretical solutions on how to control deficit when using the VCG mechanism for ridesharing problems. They show that VCG's deficit is unbounded if there are no reserve prices and propose a double reserve price scheme to control for this deficit. Altenative policies to reduce deficit are to limit the maximum detour of drivers. Agents' utilities are based on time proportional gas savings.

Herbawi and Weber (2011*c*) explores a new method to increase the probability of matching in dynamic ridesharing systems. A new bilateral trading scheme is proposed between a user who is already present in the system and assigned a ride, and a just-arrived user who could not be assigned any matched ride due to lack of available drivers. The system proposes to the two drivers a posted price based on historical data to try to exchange the ride between them. If they accept the price, the ride gets exchanged and the payment executed. System efficiency and matching probability is increased.

In other applications, VCG mechanisms have been used in procurement auctions for distributed transportation procurement (Xu and Huang, 2014). Specific dynamic primal-dual algorithms have been developed to reach the VCG outcome. These dynamic auctions reveal less information from bidders and are simpler to understand for a bidder than static VCG auctions.

This chapter fills the literature gap in pricing on P2P ridesharing systems by considering, for the first time, HOV lane time savings. The work also uses more complete utility functions that include time valuation, distance valuation and time savings due to HOV lanes. This is also the first case of the integration of a ride-back in the ride-sharing problem.

7.3 Method Preliminaries

The economic benchmark presented here is based on a mechanism design solution. Mechanism design is a field of microeconomics that studies the design of economic mechanisms. Mechanisms are methods that help a system operator to reach a particular goal, such as efficiency or revenue, while addressing potential selfish manipulation from rational economic agents who hold some information private. In our setting, we have agents who need to be assigned a role of a rider or a driver along with a trip plan, according to their private preferences. These agents hold some information private, such as about how much they value their time and cost per distance. The system operator is interested in assigning rides that maximize social welfare, while taking care of not running into a budget deficit, that is, requiring an external subsidy to implement the allocation. We provide next a short introduction to the necessary mechanism design concepts. Excellent introductions on this discipline can be found in (18, 19).

We define a mechanism M as a pair of functions $k : \Theta \to K$, $p : \Theta \to \mathbb{R}$. We call the first function the phase allocation rule that maps the type space Θ with the allocation space K. Type is the private information of the agent. The second function is the payment rule, which maps Θ to the real line, the payment space. In our notation, k can represent either the allocation function or an element in the allocation space K. We assume private information, and therefore the agents are allowed to misreport their type θ_i. Agents are assumed to have a quasilinear utility function $u_i = v_i - p_i$, in which v_i is the valuation of individual $i \in I$ and p_i is the price charged to i, which can be negative or positive.

Definition 7.1. *A mechanism M is (allocative) efficient if for any type vector $\boldsymbol{\theta}$:*

$$\sum_{i \in I} v_i(k(\boldsymbol{\theta}), \theta_i) \geq \sum_{i \in I} v_i(k', \theta_i) \quad \forall k' \in K \tag{7.1}$$

Definition 7.2. *A mechanism M is weakly budget balanced iff for any type vector* $\boldsymbol{\theta}$*:*

$$\sum_{i \in I} p_i(\boldsymbol{\theta}) \geq 0 \tag{7.2}$$

The question is now whether the sum of payments, the budget balance, is weakly positive or not. If it is not the case, the mechanism requires an external subsidy to enforce the optimal allocation. Otherwise, we say that the mechanism is financially self-sustainable.

A fundamental concept in mechanism design is truthfulness, or incentive compatibility (IC). Under IC, an agent is better off eliciting its type θ_i truthfully rather than reporting a type $\theta_i^{'} \neq \theta_i$, conditional on the rest of agents reporting truthfully.

Definition 7.3. *(Dominant-Strategy) Incentive Compatibility (DSIC)*

$$u_i(\theta_i, \theta_{-i}) \geq u_i\left(\theta_i^{'}, \theta_{-i}\right), \;\; \forall \theta_i^{'} \in \Theta_i, \; \forall i \in I \tag{7.3}$$

Where $u_i(\theta_i^{'}, \boldsymbol{\theta}_{-i})$ is agents' i utility when he reports $\theta_i^{'}$ and the rest of agents report truthfully $\boldsymbol{\theta}_{-i}$. The subindex $-i$ indicates all agents except i. Finally, we are interested in the voluntary participation of agents, that is, if users are better off participating into the mechanism than staying outside. We call this individually rationality.

Definition 7.4. *A mechanism M is ex-post individually rational for any type vector* $\boldsymbol{\theta}$*:*

$$u_i(\boldsymbol{\theta}) \geq 0 \tag{7.4}$$

We model the actual benefits of the ridesharing system with the well-known VCG mechanism. (Vickrey, 1961; Clarke, 1971; Groves, 1973). VCG is a direct revelation mechanism that is known to be efficient and (dominant-strategy) incentive compatible. VCG actually belongs to the family of Groves mechanisms, which are known to be the only ones that are efficient and (dominant-strategy) incentive compatible. Under certain conditions met in our situation and which will be elaborated later on in the text, VCG is also ex-post individually-rational. VCG is calculated in two sequential steps. First, the allocation rule which is the sum of all valuations, is maximized, and second, the prices are computed by a closed formula:

$$k^* = \text{argmax}_{k \in K} \sum_{i \in I} v_i(k) \tag{7.5}$$

$$p_i = \sum_{j \neq i \in I} v_j(k^*_{-i}) - \sum_{j \neq i \in I} v_j(k^*) \tag{7.6}$$

The price p_i of agent i is the social cost or externality that an agent exerts into the system. It is actually the difference in the sum of other agents' valuations when i is not present and when i is present. If the difference is positive, it means that agent i creates a negative externality for its presence, and thus pays a positive price. Equivalently, if the difference is negative, the price paid is negative and agent i receives a payment.

It is known that, under certain conditions, the outcome of the VCG mechanism is individually rational. These conditions are choice set monotonicity and no negative externalities. The first condition states that the choice set K expands as additional agents are introduced into the system. This is true for the ridesharing problem: adding a driver or rider will only increase the number of potential rider-driver matching options. The second property is no negative externalities. A mechanism is said to have the no-negative-externalities property when $\forall i \in I, \forall \boldsymbol{\theta} \in \Theta, \forall k^*_{-i}(\boldsymbol{\theta}_{-i}) \in K_{-i}$:

$$v_i\left(k^*_{-i}(\boldsymbol{\theta}_{-i}), \theta_i\right) \geq 0 \tag{7.7}$$

According to our definition of valuations, not considering agent i in k_{-i} leads to a valuation of user i of zero, since that agent will travel by himself along his shortest path, taken as valuation reference. Thus, we conclude that VCG is individually rational in our ridesharing setting.

It could be argued though, that an increasing presence of matched riders on HOV lanes could create a congestion externality and the former result would not hold. However, for the purpose of this chapter, the pool of participants considered is small, corresponding to the case of a ridesharing company that is starting to provide this service. In case the ridesharing service becomes successful enough to create substantial congestion, the mechanism should be embedded into an equilibration framework that includes such congestion effects. Eventually, more normal lanes should be converted to HOV lanes and provide enough supply for this beneficial service.

Finally, an interesting result of VCG following Krishna and Perry (1998) is that, in situations where VCG is ex-post individually rational, VCG is as budget-balanced as any efficient mechanism can be. In other words, no other mechanism in that situation can have a more positive budget balance than VCG. Even when compared to weaker Bayesian Incentive Compatible mechanisms. Therefore, VCG is a useful benchmark for the trade-off between budget deficit and efficiency that is examined in this chapter and we can overlook weaker implementation concepts. However, as is shown in Zhao et al. (2014), VCG's budget deficit in the ridesharing setting is unbounded. VCG will be modified by reserve prices to control for such deficit.

7.4 Matching and pricing mechanism

We present now the matching mechanism. We define two sets of agents: the drivers set D, and the riders set R. Some drivers will provide service during the morning D_M, others during the evening D_E and the rest during both peak times $D_{ME} = D_M \cap D_E$. Similarly, we can define sets R_M and R_E to include riders who wish to participate in the system in the morning and evening respectively, where $R = R_M \cup R_E$. Agents follow a quasilinear utility function as mentioned earlier: $u_i = v_i - p_i$. Users' valuations are the difference between the cost of solo driving and the cost of participating in the system, that is, the cost of being matched.

Each agent i has two private parameters to be elicited: θ_i^l, the valuation of every mile driven and θ_i^t, the valuation of every minute of the agent's travel time. The first one can be estimated from users' car model, mileage and insurance information, or alternatively, can be inputed by the users. The second parameter is user-specific and can be directly elicited or obtained by means of a short SP survey. In addition, agents submit an origin, a destination, an earliest departure time and a latest arrival time. Meanwhile, the ridesharing system operator has to provide accurate information of travel time savings, based on either the real-time state of the network, or historical data.

Drivers' valuation v_d is the negative of the detour cost in distance $\theta_d^l \delta_{d,r}^l$ and time $\theta_d^t \delta_{d,r}^t$ of driving a rider r plus the travel time cost savings ϵ_r^{HOV} in case the pair of agents can use an HOV lane. ϵ_r^{HOV} is the difference between rider's shortest path time on regular lane $SP_r^{t,REG}$ minus his shortest path on HOV lane, $SP_r^{t,HOV}$. This travel time savings depend only on the rider since the rider and driver travel through the rider's shortest path. The rider's valuation v_r is the distance cost savings of being driven through his/her shortest path SP_r^l plus the travel time savings of using an existing HOV lane. Note that the utility of not being matched is zero for both riders and drivers. Agents' valuation equations are shown next:

$$\epsilon_r^{HOV} = SP_r^{t,REG} - SP_r^{t,HOV} \tag{7.8}$$

$$v_r = \theta_r^l SP_r^l + \theta_r^t \epsilon_r^{HOV} > 0 \tag{7.9}$$

$$v_d = -\theta_d^l \delta_{d,r}^l - \theta_d^t \delta_{d,r}^t + \theta_d^t \epsilon_r^{HOV} \tag{7.10}$$

Setting the reserve prices has two functions: to reduce the budgetary deficit of the mechanism, and to classify the agents as drivers, riders or unmatched. This is particularly practical since, in day-to-day ridesharing services, the participants are often ready to play both roles: driver and rider. Reserve prices for the VCG mechanism are normally set in single-parameter environments, that is, the valuation of the agent is the actual type. In our case, our environment is multi-parameter since each agent has two private parameters. Furthermore, we have linear valuations in which each parameter type is multiplied by a variable, time or distance.

Our reserve prices are not simple scalars, but a function that maps the type and valuation to the reserve price space. If prices were just scalars for driving or being driven, there would be inefficiency due to our neglecting the fact that agents travel through routes of different lengths in both time and distance. Moreover, individual rationality and incentive compatibility could not be guaranteed. We define two constants, $\rho^t$and ρ^l, which are reserve prices for value of time (VOT) and unit distance cost VOD (value of distance), respectively. These constants are calibrated by the system operator as a function of agents' population characteristics.

We present next the VCG mechanism with multiparameter reserve prices (ρ^t, ρ^l):

1. Agents report their type $\theta_i = (\theta_i^t, \theta_i^l)$.
2. Agents are classified as drivers or riders according to:
   (a) An agent becomes a potential rider if $\theta_i^t \geq \rho^t$ and $\theta_i^l \geq \rho^l$,
   (b) An agent becomes a potential driver if $\theta_i^t < \rho^t$ and $\theta_i^l < \rho^l$.
   (c) An agent is unmatched otherwise.
3. The efficient allocation is solved.
4. VCG prices p_i^{VCG} are calculated.

5. Agents are charged the following prices:

$$p_r = \max\{\rho^l SP_r^l + \rho^t \epsilon^{HOV}, p_r^{VCG}\}$$

$$p_d = \max\{-\rho^l \delta_d^l - \rho^t \delta_d^t + \rho^t \min\left\{\delta_d^t, \epsilon^{HOV}\right\}, p_d^{VCG}\}$$

Essentially, the purpose of the above reserve price is for the riders to be charged at least the maximum payment that the limit type rider $(\theta_i^l, \theta_i^t) = (\rho^l, \rho^t)$ can pay such that this agent's individual rationality is respected. This guarantees a larger positive revenue from the drivers. On the other hand, the reserve price for drivers ensures that no driver is compensated more than the minimum compensation for the binding type driver. Thus, the payments that drivers receive are minimized while respecting still their individual rationality constraints.

The main reason why this mechanism has a lower deficit than the VCG mechanism is because the reserve prices filter agents such that higher-valuation agents are the riders who make positive payments, and the lower-valuation agents are the drivers who receive payments. Thus, positive payments are increased and negative payments are minimized. Furthermore, the travel time savings due to HOV lanes has a positive effect on the budget. HOV-related payments from riders is always larger than those received by the drivers since the former are larger than the latter, while ϵ^{HOV} is the same for both agents.

Proposition 7.1. *The mechanism is incentive compatible and individually rational.*

Proof. Let i be a rider. If $p_i^{VCG} > \rho^l SP_i^l + \rho^t \epsilon^{HOV}$, his individually rationality is guaranteed by the argument made earlier on the standard VCG mechanism. Else, since $\theta_i^t < \rho^t$ and $\theta_i^l < \rho^l$ the reserve price charged is always smaller than the valuation. Analogous is the case in which i is selected as a driver. Supose that agent i is classified as a driver. If the driver misreports a higher type, he/she will be classified as a rider and charged either a reserve price that will violate his/her individual rationality, or will be charged a VCG payment that will not maximize his/her utility, since VCG is incentive compatible. Analogously, if the agent is classified as a rider, by misreporting a lower type his/her utility will not be maximized. □

To ensure truthfulness, it is important that the optimization of the allocation rule is exact, not approximate. Otherwise, truthfulness cannot be guaranteed (Nisan and Ronen, 2007). Our problem formulation is in a min-cost max-flow form, which has exact integer solution since the constraint matrix is totally unimodular. We guarantee that the solution is optimal by using standard linear programming techniques.

An aspect left aside so far are the users who do not own a vehicle or do not have temporary access to one. From the mechanism definition, users who do not own a car would mostly fall into the category of drivers, since they would have to specify a VOD of either zero or very low one in case their default mode of transportation is public transit. A second similar case arises, which is that of users who own a car but they do not have it temporarily available and would still wish to share a ride, such as for instance, due to mechanical failure or because a family member had to use that car.

An idea to accommodate these two types of users is to set up a “leftover pool” that takes advantage of the budget surplus resulting from the main optimization and provides these agents with a ride from the unmatched drivers or through an external service, when possible or desirable, as is often done in carpooling companies. Obviously, measures need to be established to prevent users from manipulating and gaming the system. Some examples are (1) providing a justification for not having the car available, (2) setting up a maximum number of leftover pool rides for each user and (3) setting up a penalty for lying about having a car available. For users who do not own a car, a separate, more expensive fare could be set, since they do not contribute to the driver supply. Naturally, priority in the leftover pool will be given to the users with a history of participating in the system with a vehicle.

We present the allocation rule in a min-cost max-flow formulation, as in Figure 7.1. Once agents are classified as drivers and riders, they are grouped in columns. Assuming a daytime problem, with ride-backs being in the evening (as opposed to night-shift travel), the first column consists of the morning drivers, and the second column, of the morning riders. The riders are duplicated for the evening matching, and finally there is the evening driver column. Each matching arc is assigned a cost $c_{ij} \in \{0, c_{dr} = -(v_r(r,d) + v_d(d,r))\}$. Rider morning to rider evening arcs have cost equal to zero.

At the initial extreme, there is a source vertex, which generates $b_s = \min\{|D_M|, |R|, |D_E|\}$ units of flow. At the other end, there is a sink vertex that attracts $b_t = -b_s$ units of flow. Source-to-driver and driver-to-sink arcs have costs equal to zero. In order to make the allocation individually rational and as efficient as possible, we add a source-to-sink arc with zero cost and capacity equal to b_s. This link prevents the existence of negative welfare driver-rider-driver matchings. The rest of arcs have their capacity set to one.

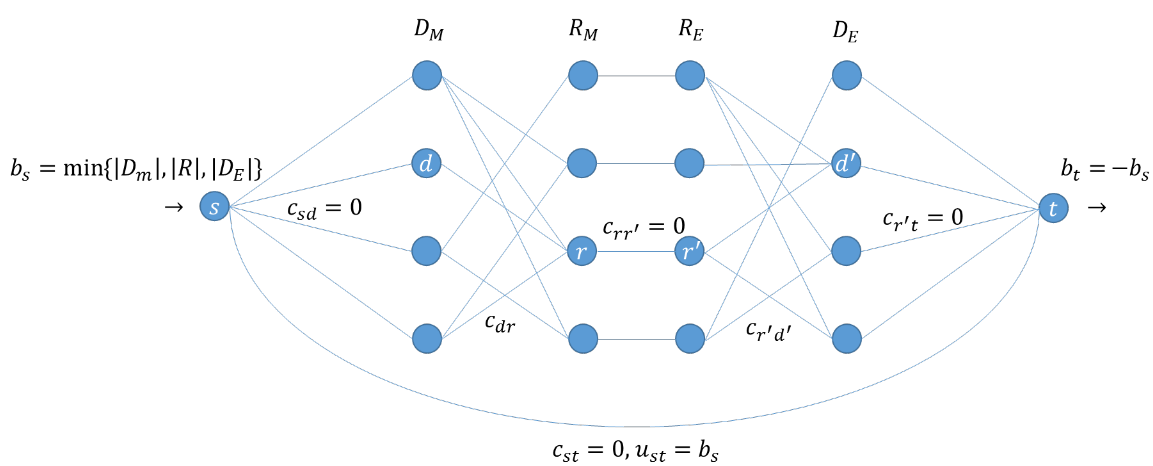


Figure 7.1: Min-cost max-flow graph for the ridesharing with ride back problem

The above graph is built after executing a preprocessing step in which agents that do not intersect on space and time do not have an arc connecting them. This considerably reduces the size of the graph and therefore the optimization time. The VCG prices for agents i are calculated by subtracting the edges that connect to agent i vertices and solving the optimization again. By using the former solution as initial solution, the optimization speed is increased considerably. Please note that there is always a solution possible, that diverts all the flow through the source-to-sink auxiliary arc.

The min-cost max flow formulation can be equivalently represented as the following linear program:

$$\min \sum_{(i,j)\in A} -c_{ij}x_{ij} \tag{7.11}$$

$s.t.$

$$\sum_{(i,j)\in\delta_i^+} x_{ij} - \sum_{(j,i)\in\delta_i^-} x_{ji} = b_i \tag{7.12}$$

$$b_i = \begin{cases} \min\{|D_M|,|R|,|D_E|\}, & i = s \\ -\min\{|D_M|,|R|,|D_E|\}, & i = t \\ 0, & else \end{cases} \tag{7.13}$$

$$u_{ij} = 1, \forall\,(i,j) \in A - \{(s,t)\} \quad u_{ij} = b_s, \forall\,(i,j) = \{(s,t)\} \tag{7.14}$$

$$0 \le x_{ij} \le u_{ij},\ \forall\,(i,j) \in A \tag{7.15}$$

Equation (7.11) is the objective function, the sum of the negative of all edge costs times the edge flows x_{ij}. Constraints (7.12) represent the flow conservation at every node, where (7.13) defines nodes' supply terms b_i. $\delta_i^+$and $\delta_i^-$are the set of outgoing and incoming edges from node i. Constraints (7.14) and (7.15) define edge capacities u_{ij}, according to the discussion made earlier for the graph.

7.5 Data sources and treatment

This chapter relies on five data sources. Origin and destination nodes and their AM and PM peak demands for a 3-hour period are extracted from the Southern California Association of Goverment (SCAG) model of 2012. In total, we consider 3051 nodes corresponding to 9,308,601 OD pairs. Our study area, shown in Figure 7.2, covers Los Angeles County and parts of San Bernardino County and Orange County which encapsulate the totality of Southern California HOV network. Since the SCAG model is a planning model, the trip tables from this model are only available at an aggregate level. For the purpose of our study, we distribute the peak-hour demand throughout the duration of morning and evening peak hours based on a uniform random distribution.

OD travel times and distances are obtained by using the Google Maps API. The following information was requested at 8AM and 5PM for Wednesday July 20^{th} (a typical work day): total travel time given traffic, total trip distance, total freeway distance if an HOV is present. The presence of HOV lanes was determined by parsing the current HOV-equipped freeways from the retrieved trip instructions. Querying travel times from Google Maps is a costly task in time and money, since there are quotas for their usage. For this reason, the nodes were aggregated geographically to reduce the number of OD queries. This was done by means of a k-means clustering procedure. The number of clusters was set to 136 which led to small circular areas of 2.5 km of radius on average, with an average max radius of 6km. Figure 7.2 shows the study area with the nodes, cluster centroids and (average and maximum) cluster areas.

Final ODs distances and times are computed with some approximations. If the distance between the origin and the destination is shorter than 3 km or if both nodes belong to the same cluster, the OD trajectory is Euclidean and the speed is set to 25mph (40kmh). Otherwise, the ODs are composed by three segments: origin node to the origin cluster centroid, the origin cluster centroid to the destination cluster centroid, and a final segment from the destination cluster centroid to the destination node. This extra distances are run at 25mph (40kmh) as well.

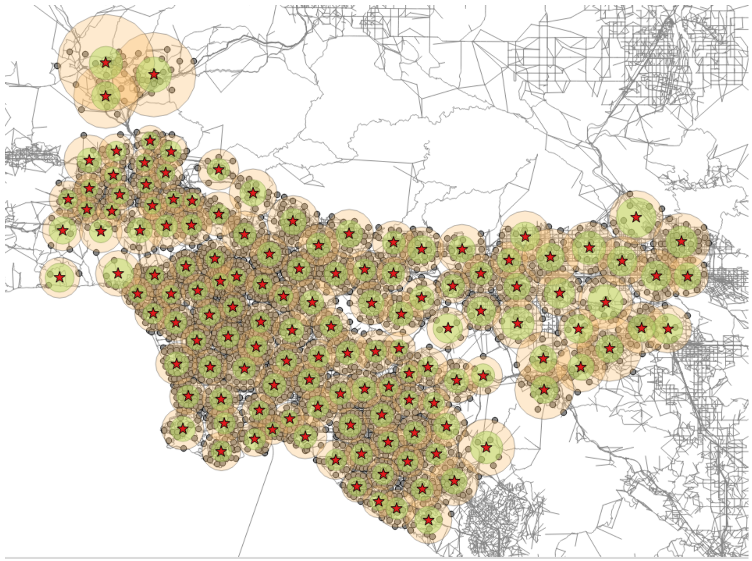


Figure 7.2: Study area showing nodes, cluster centroids (star) and cluster areas (mean radius in green, maximum radius in pink).

Google Maps does not provide HOV travel times. To include the HOV travel time savings, we obtained the relationship between mainlane speed and HOV speed from NGSIM data collected at I-80 (FHWA, 2016). This data covers the period from 5:00PM to 5:30PM. We extracted the average speed for the regular lanes (11 mph), the average speed for the HOV lane (27 mph) and established an affine relationship between the HOV lane speed s_{HOV}, and the regular lanes' speed s_{ML}, given the free flow speed s_{FF} of 65 mph. This relationship provides the HOV speed given the regular lane speed obtained from Google maps, after having subtracted the above mentioned non-freeway segments of the path.

$$s_{HOV} = s_{FF} - \frac{s_{FF} - 27}{s_{FF} - 11}(s_{FF} - s_{ML}) \tag{7.16}$$

With more extensive data on HOV lane speeds, the system could propose optimal departure and pickup times for the rider and driver and reach further time savings. The same relationship is used for both peak periods.

We acknowledge that some degree of positive correlation could exist between VOT and VOD. Higher income people are observed to value time more and to purchase vehicles with higher operating costs (Small and Verhoef, 2007). But given the complexity of vehicle purchase decision and that the VOT increases with income less than proportionally (Mackie et al., 2003), we assume independence between VOT and VOD. Thus, their values will be drawn from two independent distributions. Agents' VOT are drawn from the VOT distribution estimated in Abou-Zeid et al. (2010) that used a Swedish SP survey based on car alternatives, accounting for income levels and latent variables.

Driving costs per distance are obtained from AAA's Your Driving Costs study for year 2014 (AAA, 2014). This study gives the costs per mile for the following vehicle segments: small, medium and large sedan, SUV 4WD and minivan. This costs include fuel, maintenance, tire, insurance and depreciation costs. Since we assume that agents keep their own vehicle, we discount insurance and depreciation costs for the driving costs. These latter items account for around 60% of of the costs. The VOD distribution is obtained by weighting the former costs by a distribution of vehicle segments. The latter is obtained from the car sales in 2016 from January to June 2016 (The Wall Street Journal, 2016).

7.6 Simulation Results

This section presents the simulation results of our control policy on the study area decribed above. The total number of agents is always set to 2500 and six different seeds are used for pseudo-random population characteristics generation.

7.6.1 Percent of drivers, riders and supply to demand ratio

We first observe the effect that the reserve prices exert on the induced supply and demand of the ridesharing services. Please note that the percentages of drivers and riders do not add up to a hundred since those vehicles that do not satisfy the reserve prices are left unmatched and drive by themselves. Figure 7.3 shows how increasing ρ^t and ρ^l has a positive effect on the percent of drivers and a negative effect on the quantity of riders. This leads to the riders-to-drivers ratio increasing as we decrease the prices. We thus expect higher social welfare on the lower reserve price section. From the graph on the right, we can observe that the interesting regions for our sensitivity analysis are the 10 to 30 $/h range for the VOT and the 0.35 to 0.6 $/mile range for VOD (value of distance).

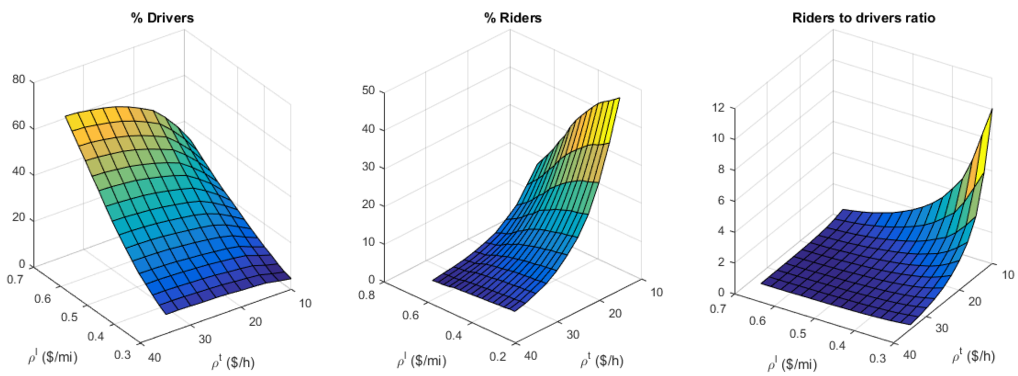


Figure 7.3: Percentages of drivers and riders, and ratio of riders to drivers as a function of the reserve prices.

7.6.2 Sensitivity analysis over both reserve price components.

We next analyze the sensitivity of the matching rate in our model to the reserve prices, setting the maximum detour for the drivers to 20 minutes. Figure 7.4 shows the revenue per vehicle, total social welfare SW (the sum of all agents' utilities) and the percentage of matched participants.

The percentage of matched participants decreases with ρ^t and increases with ρ^l. While

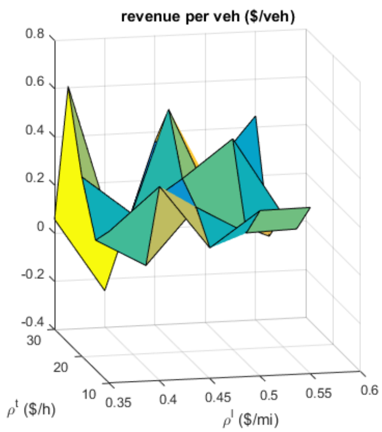


Figure 7.4: Revenue per vehicle, total social welfare and percent of matched participants.

the former effect is expected, since the ratio of riders to drivers increases from below one to more than, the latter requires more attention. The main explanation is that by raising the reserve price on distance, the agents who become riders have higher cost savings and higher probability of being matched. This eventually compensates the decline in matching due to the decreasing ratio of riders to drivers. An observation worth making is that a high percentage of vehicles getting matched could increase congestion in the HOV lane and reduce its travel time benefits. However, which such a low number of participants simulated, this effect is considered to have little impact. This could be the case of the initial states of the implementation of the ridesharing policy, in which the system has not gained a critical mass of participants.

From the center plot, we observe that the participant's total social welfare is correlated with the percentage of matched participants. From this sensitivity analysis, we conclude that the most favorable reserve prices are low ρ^t and high ρ^l. That is, around 10$/h and 0.55$/mi respectively. The leftmost figure shows how the average revenue per vehicle stays positive over all the reserve price parametric domain, except for the corner with high ρ^t and low ρ^l, which is of little operational interest. Over the low ρ^t, high ρ^l region, the revenue oscillates between 18 and 38 cents per vehicle. The instability of the revenue surface is explained by the non-convexity of the reserve price expressions.

7.6.3 Sensitivity analysis on maximum driver detour and VOT reserve price.

We analyze now the consequences of limiting the maximum detour length on drivers. Maximum driver detour is an operational parameter that the system manager can use to control for revenue and efficiency. Maximum detour has two effects on the system: increasing it may increase the total number of matches and therefore the social welfare, but it also increases the deficit, since drivers need to be compensated for longer detours while riders keep paying similar fares.

To be able to analyze the influence of the maximum detour, we fix the VOD reserve price to 0.54 cents per mile, which is the Standard IRS (Internal Revenue Service, USA) mileage rate for 2016. This mileage rate accounts for fixed and variable costs for operating a vehicle. Moreover, this mileage rate turned out to be the maximum social welfare point in the previous section. We enrich the detour analysis with sensitivity on the VOT reserve price.

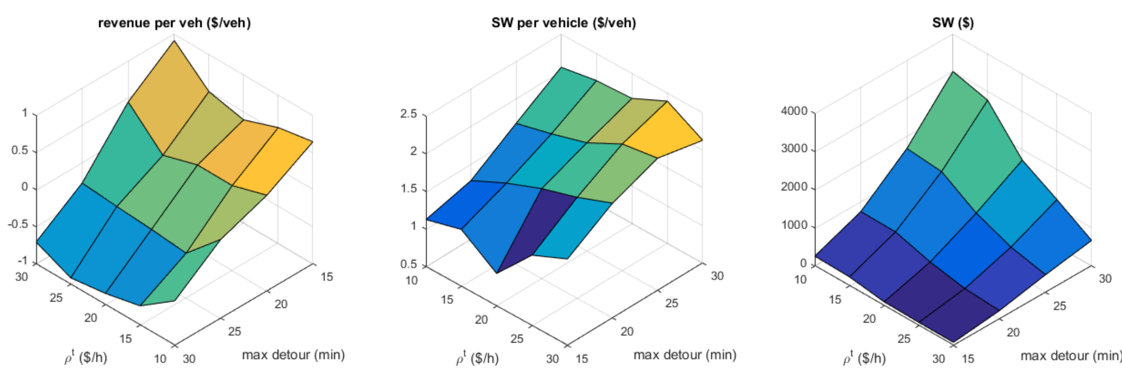


Figure 7.5: Revenue per vehicle, Social welfare per agent, Total social welfare per vehicle.

Leftmost plot in Figure 7.5 shows how the revenue per vehicle decreases as the maximum detour increases, leading to negative revenues above the 25 min of maximum detour. The reserve price on the VOT seems to slightly increase revenue, due to higher savings from HOV lane per vehicle and also for reducing the actual number of agents matched, as seen in Figure 7.4 fewer matches are made, but more revenue per match is collected. Similarly, increasing the detour increases the average utility per vehicle and total social welfare since more vehicles are matched. Finally, increasing the reserve price on VOT has a positive effect on the social welfare per vehicle but a negative one on total social welfare. An explanation of the latter may be due to the increase in revenue mentioned earlier.

Finally, we analyze the effect the maximum detour has on the percent of riders matched with the same driver for both morning and evening tour. The results are shown in Figure 7.6. We observe how an increase on the maximum detour time for drivers translates into more drivers being matched with the same driver for both trips. As drivers' travel time budgets become longer, the set of potential matches per rider increases and the most efficient one is selected. If the most efficient driver is selected during the morning for a rider, that same driver should be selected during the evening as well.

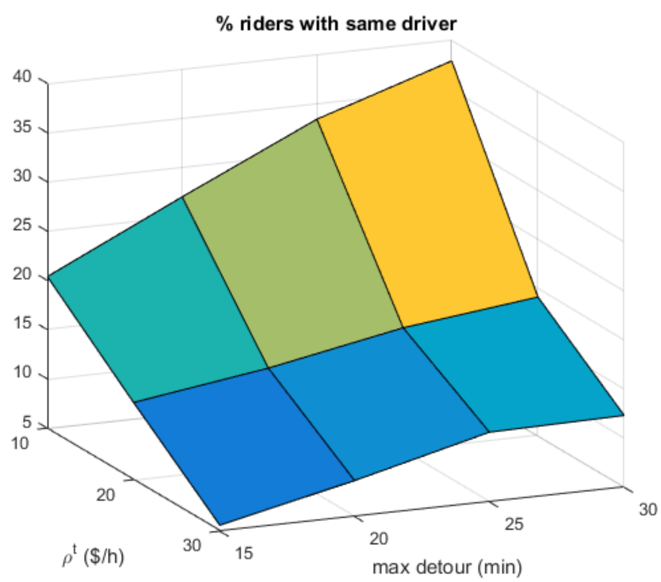


Figure 7.6: Percent of matched riders with the same driver for both trips

This trend is present over the whole reserve price range, but it is clearer for lower VOT reserve prices since the riders-to-drivers ratio increases. From these results, we conclude that modeling the ride-back separately is generally beneficial for a system operator interested in social welfare, while for a revenue oriented operator, bundling the morning and evening ride into a round trip may help increase the revenue.

7.7 Conclusion

We presented an economic benchmark of the interaction of P2P ridesharing with HOV lanes. Ridesharing was introduced as an alternative mode for commuting that can provide not only cost savings via the sharing of trips, but also time savings due to the possibility of HOV lane use. The ridesharing service, modeled as an economic mechanism, is based on the VCG mechanism which is known to be efficient, incentive compatible and individually rational in the ridesharing environment. However, VCG is known to run budget deficits, making the system financially unsustainable without external subsidy.

A multiparameter reserve price was introduced to control for the budget deficit. In addition, it efficiently classifies agents between drivers and riders before the matching, and so eliminates the need to exogenously specify who is a driver and who is a rider. The reserve price successfully manages to collect enough revenue for most of the parameter space, making the system financially sustainable. This is the case of low reserve price for the time cost and high reserve price for the distance cost, the combination for which the sum of agents' utility is maximized. Finally, we conclude that modeling the ride-back as a separate complementary trip increases social welfare and the percentage of matched participants.

We suggest next several lines of further research to extend this work. First, extending the reserve price method to more complicated ridesharing settings such as multi-hop day-to-day ridesharing. This would require a careful examination into satisfying incentive compatibility, since it is not guaranteed when the allocation rule is not exact. Second, using richer utility functions that could include extra time savings, such as parking savings for the rider. Third, modeling the congestion resulting from the interaction between matching percentage and HOV travel time benefits. This could be addressed by embedding the mechanism into an equilibration framework. Moreover, planning and simulation models that have sharing economy inherent in them should be developed. Fourth, adding a day-to-day choice-related behavior component in agents would provide a better understanding of the actual ridership volume of this system.

Chapter 8

Conclusion

Consumption of supply in transportation systems has always followed an FCFS service order. This policy is inefficient and unfair when user heterogeneity is considered into the system. With the advent of smartphones and connected vehicle technologies, it is possible to incorporate real-time information in the operation of transportation systems.

This dissertation presents a new paradigm, Collaborative Consumption of Transportation Supply, which violates and outperforms FCFS not only in efficiency, but also in the fairness dimension. Collaborative Consumption consists of decentralizing the consumption of supply to the users, by allowing them to execute real-time trades, either directly or through a broker. In the former case, I call the policy P2P Exchange of supply.

This research has (1) presented a new paradigm in the operation of transportation systems, both conceptually, analytically and experimentally, (2) created new methods to make this paradigm work, such as Dynamic Envy-freeness, and (3) explored this paradigm on several novel applications in different transportation areas: exchange-based traffic signal control, dynamic P2P ridesharing and day-to-day ridesharing and queue-jumping operations.

8.1 Summary of contributions

8.1.1 Supply and Demand framework for Collaborative Consumption of Supply

To present this newer paradigm of operation with collaborative consumption of transportation supply within which the non-FCFS schemes are explored, I first developed a conceptual theoretical supply-demand framework, with appropriate modifications from the classic and traditional view of such frameworks in transportation theory. The emphasis is on incorporating individual pricing and information exchange into the traditional frameworks. While it is conceptual in nature and is not a solution to the problem, the enhanced framework is useful for giving an appropriate fundamental structure to the dissertation.

8.1.2 Concepts based on Envy for Pricing and Behavior, and extensions to Dynamic Envy-freeness and CESEI criteria

Envy-free pricing has been presented as a method that has proved useful for Collaborative Consumption of supply. This pricing policy intrinsically represents an alternative notion of fairness which can substitute the historical status quo of FCFS. An allocation is envy-free when no agent is better off with some other agent's allocation at the current prices than with his own. In addition, envy-freeness has been extended to the time dimension, creating Dynamic Envy-freeness. Though not strongly presented as a behavior mechanism, the dissertation alludes to the relevance of the concept of envy for behavioral modeling of user response and decisions where tradiitional models in transportation are rather inadequate, owing to their normal assumption of users lack of awareness of other users's data on an individual basis. Thus the concept of envy can form a basis for both pricing from an optimization standpoint, and behavior from a modeling standpoint, when the contexts involves exchanges among users.

Secondly, I extended the theory of relaxed envy-freeness. I designed a new envy-minimization criteria family, the Constant Elasticity of Substitution Envy Intensity (CESEI) criteria family. CESEI provides a new normative interpretation of envy-freeness, in which users' envy is no longer understood as the minimal excess envy such that they don't feel envy with the most envied agent, but as polynomial combination of envy intensities. CESEI is shown to be a flexible criterion which outperforms existing criteria in the literature. This criteria is successfully tested on the proposed traffic signal control PEXIC.

8.1.3 Priced EXchanges in Intersection Control

I designed a new traffic signal control scheme which takes into account user heterogeneity in value of delay savings. The method presumes that current and imminent technology renders data on such individual valuations available for control. PEXIC allows for signal timings to be determined without insisting on the traditional first-come first-served (FCFS) order, which we believe can be broken if fairness in service can be preserved under utility exchanges via payments among users.

PEXIC successfully manages to increase efficiency and fairness for a large range of critical volumes and users' heterogeneity, regardless of the actual type of user. PEXIC introduces a novel envy-minimizing (fair) pricing approach which compensates for the additional delay differences between agents by means of utility transfers. Fairness is addressed by minimizing the envy perceived by users about the monetary transfers and experienced delay of others. Furthermore, the control mechanism is Pareto-efficient, budget-balanced, and therefore it does not require any external subsidy to operate.

8.1.4 P2P Ride Exchange in Dynamic Ridesharing Systems

On the totally different field of ridesharing operations, I proposed P2P Ride Exchange. The idea is summarized next: a user who is in a hurry joins the system but does not find a ride. Another user who joined the system earlier, is waiting for a ride and also has a second-best ride. P2P Ride Exchange consists of transferring ride property rights to users in such a way that they can trade their rides if both can be better off.

P2P Exchange of supply successfully increased the efficiency levels of a dynamic P2P ridesharing system. The posted-price mechanism used is robust and has low information requirements. In addition, higher performance levels and customer retention rates are achieved when compared to FCFS allocation. Numerical experiments show that the percentage of increase in the matching rate using P2P ride exchange is positively correlated with the number of participants, travel time budget of riders and drivers, and spatial and temporal proximity of trips. Finally, the exchange mechanism is more effective when the ratio of number of riders to system participants is medium.

8.1.5 Day-to-day P2P Ridesharing systems

In the last application of this dissertation, I used market design tools to make P2P ridesharing an attractive mode for daily commuting. Ridesharing was introduced as an alternative mode for commuting that can provide not only cost savings but also time savings due to HOV lane savings. The ridesharing mechanism is a multiparameter reserve price VCG mechanism which is truthful and financially self-sustainable. Moreover, it efficiently classifies agents between drivers and riders before the matching, and thus eliminates the need to exogenously specify who the driver is, and who the rider. Modeling the ride-back as a separate complementary trip increases social welfare and the percentage of participants matched.

8.1.6 Traffic operations: cooperative dynamic routing and queue-jumping operations

First, I presented a new cooperative dynamic routing scheme applicable for freeway control and facilities with parallel queues. This control scheme is seen to be core-stable for a vertical queue modeling and very stable for the dynamic queue modeling. The control algorithm is based on a partition function game, which I believe is the first application of this kind of techniques in transportation literature. This game incorporates the both the strategic and coalitional interaction of agents. A new relaxation for the strong-core for partition function games is presented.

Second, I introduced a new concept: queue-jumping for general freeway operations. This policy allows vehicles which are in a hurry to overtake vehicles already present in a queue by paying them, such as the resulting queue is stable and fair. Queue-jumping operations take advantage of automated and connected vehicles technology. The proposed algorithm is Pareto efficient and fairer than FCFS (envy-free).

8.1.7 Final remarks on contributions

In conclusion, the contributions mentioned above validate P2P exchange and collaborative consumption of supply as a new paradigm operational paradigm for transportation science, which overcomes FCFS in both efficiency and fairness. I have also examined other objectives such as truthfulness and coalitional stability. The methods employed and created in this dissertation can prove as useful benchmarks for this new paradigm, not only for more detailed versions of the examined applications but for different, new situations which may arise in other modes of transportation.

8.2 Further research

The methodological focus of this dissertation was to ensure the efficiency and fairness for collaborative consumption of supply. The former concerns social welfare, while the latter one addresses public acceptance. However, there are other dimensions worth exploring, such as incentives. While incentives were indeed explored in the two ridesharing applications, either as a posted-price mechanism or as a multiparameter reserve price VCG, it was not the case for the traffic operations.

Still, I do not think that incentives will represent a great operational problem in the traffic operations, at least as far as Incentive Compatibility is concerned, as pointed out in the relevant PEXIC subsection. Other cases where incentives can matter are collusion and shill bidding. However, I would qualify these issues as security threats rather than an economic incentives issue. In the end, these strategic issues could be prevented by proper detection, identification and participation control measures.

8.2.1 Methods

The version of dynamic envy-freeness explored in this dissertation was thought for a rolling horizon optimization. Thus, it is discrete in time. However, other specifications could be designed. For instance, the discrete points of envy measurement could span the time dimension and not just rolling horizon periods. Also, the whole envy specification could be on the continuous domain and include more complex utility specifications, such as budgeted agents, risk aversion, loss aversion, etc.

The development of Constant Elasticity of Substitution Envy Intensity criteria family has created a new interpretation of envy. The envy of an agent is no longer the maximum of the pairwise envies he or she suffers but an aggregation of all the pairwise comparisons.

This new interpretation opens the door to microeconomic explanations which can justify these new measures. Furthermore, CESEI criteria excess envy could be experimentally calibrated with humans, and we can further enrich the criteria with agents' income or socio-economic characteristics such as sex, age or mode of transportation.

8.2.2 Depth: multiple operators and network-level control

The exposition of the supply-demand framework presented two new directions to be tackled in future work: collaborative consumption with multiple operators and interaction of demand with the decentralization of supply. While operators can be modeled as agents, as I have done with the infrastructure manager, there is a new dimension now to be considered: privacy preservation. Operators, either in competition with other operators or under supervision of a regulator, may not be willing to share many of the information necessary to effectively develop or monitor collaborative consumption. This unwillingness may be motivated by profit maximization or reluctance to reveal business secrets, and it may directly concern both users and regulators. Thus, effective privacy-preserving mechanisms and communication protocols must be explored for collaborative consumption with (multiple) private operators.

So far, the traffic control policies presented in this dissertation were of local impact. It is worth exploring how these policies can be adapted to the network level. In particular, for PEXIC, the policy needs to be evaluated on network-level simulations to assess the efficiency scalability. Users exchanging supply in real time may induce changes in OD patterns and individual budget imbalances. An interesting possibility is the directional effects in urban networks, such as when drivers with higher urgency may be driving in the radial directions to CBDs, while drivers of lower urgency may be more prevalent in directions orthogonal to them. This also raises possibilities of flow control on routes on the basis of subsidies (or credits) that would channel flows towards more efficient use of network capacity. Concerning queue-jumping operations, it is worth investigating the resulting efficiency levels with real car-following models to gauge the effects of gap creation and participant set creation.

8.2.3 Scope: Collaborative Consumption in Urban Systems

Finally, the methods and applications explored in this dissertation can be extended to what I call Collaborative Consumption in Urban Systems, summarized in Figure 8.1. Urban systems are composed of many interrelated systems which concern different activities developed within the urban environment. In this dissertation, I addressed the transportation activity layer, in both private transportation (traffic signal control and highway operations) and public transportation (P2P dynamic ridesharing and day-to-day ridesharing) sides.

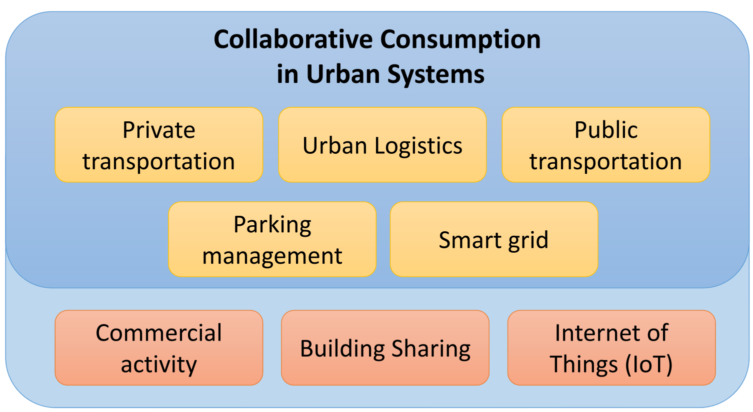


Figure 8.1: Collaborative Consumption in Urban Systems

Transportation services are generally a derived demand from other activities, i.e. work, commercial, business or leisure. Thus, the use of urban infrastructure supply is obviously related these activities. These activities can be similarly included to the supply-demand framework presented in this dissertation and included in agents' utility specifications a alternative (trade) choice sets.

I provide next some examples of such extensions. The most straightforward one is the inclusion of urban logistics into the traffic network. For instance PEXIC and queue-jumping operations could include urban logistics vehicles either directly through the valuation elicitation protocol or creating dedicated corridors for these activities in order to maximize their efficiency and added value.

In the same way that new connectivity possibilities justified the study of collaborative consumption of transportation supply, other innovative technologies such as automated and electric vehicles allow new operational schemes. This is the case for parking management and smart grid. (Automated) Electric vehicles present new opportunities and solutions for problems such as peak electricity demand, empty ride-backs for self-driving vehicles and parking search problems.

Finally, collaborative consumption of supply could be expanded to building sharing, i.e. remote attendance of classes, telecommuting, shared usage of public spaces. Also Internet of Things (IoT) brings new capabilities in urban network management, i.e. dynamic value congestion toll for urban logistics tours, where truck load information is gathered with sensors, or, similarly, for fleet management in Urban Consolidation Centers.

8.3 Final Comments

This dissertation is the result of a journey through ideas and concepts, some of which initially appeared outlandish, but later proved to be more and more productive as the research progressed. It is not just humility that makes me say that it is by no means a complete and comprehensive study of everything I proposed; rather it is just the realization that every box opened during the work was only showing further boxes inside to open, like opening Russian Matryoshka dolls and finding further dolls inside. Except that in this cases the dolls for further play may not be smaller every time. Indeed, issues for related studies in the future could rather be larger. In that spirit, I must direct the reader back to the introductory chapter's final words, on what the dissertation set out to do and what it did not. Much further work remains to be done, but that is what pushes forth the academic endeavor.

Bibliography

AAA (2014), 'AAA Your Driving Costs 2014'.
URL: *www.newsroom.aaa.com/tag/your-driving-costs-2014/*

Abe, T. (2016), 'Efficiency and the Core in Games with Positive and Negative Externalities', *WINPEC Working Paper Series* (April).

Abou-Zeid, M., Ben-Akiva, M., Bierlaire, M., Choudhury, C. and Hess, S. (2010), Attitudes and value of time heterogeneity, *in* 'Applied Transport Economics-A Management and Policy Perspective. De Boeck Publishing', pp. 523–545.

Agatz, N. a. H., Erera, A. L., Savelsbergh, M. W. P. and Wang, X. (2011), 'Dynamic ride-sharing: A simulation study in metro Atlanta', *Transportation Research Part B: Methodological* **45** '(9), 1450–1464.

Agatz, N., Erera, A., Savelsbergh, M. and Wang, X. (2010), 'Sustainable Passenger Transportation: Dynamic Ride-Sharing', *ERIM Report Series Reference No. ERS-2010-010-LIS* .

Alkan, A., Demange, G. and Gale, D. (1991), 'Fair Allocation of Indivisible Goods and Criteria of Justice', *Econometrica* **59**(4), 1023–1039.

Alt, H., Blum, N., Mehlhorn, K. and Paul, M. (1991), 'Computing a maximum cardinality matching in a bipartite graph in time O (n 1.5 mlog n)', *Information Processing Letters* **37**(4), 237–240.

Arnott, R., de Palma, A. and Lindsey, R. (1990), 'Economics of a bottleneck', *Journal of Urban Economics* **27**(1), 111–130.

Baldacci, R., Maniezzo, V. and Mingozzi, A. (2004), 'An Exact Method for the Car Pooling Problem Based on Lagrangean Column Generation', *Operations Research* **52**(3), pp. 422–439.

Belk, R. (2014), 'You are what you can access: Sharing and collaborative consumption online', *Journal of Business Research* **67**(8), 1595–1600.

Bento, A., Roth, K. and Waxman, A. (2015), 'The Value of Urgency : Evidence from Congestion Pricing Experiments', *Working paper* .

Bialas, W. F. (1989), 'Cooperative n-person Stackelberg games', *Proceedings of the 28th IEEE Conference on Decision and Control* **3**, 2439–2444.

Bialas, W. F. and Chew, M. N. (1982), Coalition Formation in n-Person Stackelberg Games, *in* '21st IEEE Conference on Decision and Control', pp. 1–5.

Borm, P., Hamers, H. and Hendrickx, R. (2001), 'Operations Research Games: A Survey', *TOP* **9**(2), 139–199.

Bradford, R. M. (1996), 'Pricing, routing, and incentive compatibility in multiserver queues', *European Journal of Operational Research* **89**, 226–236.

Brams, S. J. and Taylor, A. D. (1996), *Fair Division: From cake-cutting to dispute resolution*, Cambridge University Press.

Brereton, M. and Ghelawat, S. (2010), Designing for participation in local social ridesharing networks: grass roots prototyping of IT systems, *in* 'Proceedings of the 11th Biennial Participatory Design Conference', ACM, pp. 199–202.

Buttle, F. A. (1998), 'Word of mouth: understanding and managing referral marketing', *Journal of strategic marketing* **6**(3), 241–254.

Careerbuilder.ca (2015), 'One Third of Canadian Workers Live Paycheck to Paycheck, CareerBuilder.ca Survey Finds'.
URL: *careerbuilder.ca/ca/share/aboutus/pressreleasesdetail.aspx?sd=2/19/2015&id=pr72&ed=12/31/2015*

Carlino, D., Boyles, S. D. and Stone, P. (2013), 'Auction-based autonomous intersection management', *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC* (October), 529–534.

Cervero, R. and Griesenbeck, B. (1988), 'Factors influencing commuting choices in suburban labor markets: a case analysis of Pleasanton, California', *Transportation Research Part A: General* **22**(3), 151–161.

Chan, N. D. and Shaheen, S. A. (2012), 'Ridesharing in north america: Past, present, and future', *Transport Reviews* **32**(1), 93–112.

Chander, P. (2014), 'A Core Concept for Partition Function Games', *Working paper, Jindal School of Government and Public Policy* .

Chander, P. and Tulkens, H. (1995), 'A core-theoretic solution for the design of cooperative agreements on transfrontier pollution', *International Tax and Public Finance* **2**(2), 279–293.

Chander, P. and Tulkens, H. (1997), 'The Core of an Economy with Multilateral Environmental Externalities', *International Journal of Game Theory* **26**(3), 379–401.

Chaudhuri, A. (1986), 'Some implications of an intensity measure of envy', *Social Choice and Welfare* **3**(4), 255–270.

Chevaleyre, Y., Endriss, U., Estivie, S. and Maudet, N. (2007), 'Reaching envy-free states in distributed negotiation settings', *IJCAI International Joint Conference on Artificial Intelligence* pp. 1239–1244.

Chun, Y. (2016), *Fair Queueing*, Studies in Choice and Welfare, Springer International Publishing.

Chun, Y. and Jeong Heo, E. (2008), 'Queueing problems with two parallel servers', *International Journal of Economic Theory* **4**(2), 299–315.

Clarke, E. H. (1971), 'Multipart pricing of public goods', *Public Choice* **11**(1), 17–33.

Cleveland, W., Grosse, E., Shyu, W., Chambers, J. and Hastie, T. (1992), Local regression models, *in* 'Statistical Models in S', pp. 309–376.

Coase, R. (1960), 'The Problem of Social Cost', *The Journal of Law & Economics* **3**(10), 1–44.

de Clippel, G. and Serrano, R. (2008), 'Bargaining, Coalitions and Externalities: A Comment on Maskin', *IMDEA Ciencias Sociales: Working papers series in Economics and Social Sciences* pp. 1–14.

Di Febbraro, A., Gattorna, E. and Sacco, N. (2013), 'Optimizing Dynamic Ride-Sharing Systems', *Proceedings of 92th Transportation Research Board* .

Diamantaras, D. and Thomson, W. (1990), 'A refinement and extension of the no-envy concept', *Economics Letters* **33**(3), 217–222.

Do, K. H. P. and Norde, H. (2007), 'The Shapley Value for Partition Function Form Games', *International Game Theory Review* **09**(02), 353–360.

Dresner, K. and Stone, P. (2004), 'Multiagent traffic management: a reservation-based intersection control mechanism', *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems, 2004. AAMAS 2004.* (July), 530–537.

Endriss, U., Maudet, N., Sadri, F. and Toni, F. (2006), 'Negotiating socially optimal allocations of resources', *Journal of Artificial Intelligence Research* **25**, 315–348.

Enrico, H., Robu, V., Stein, S., David, C., Rogers, A., Online, N. R. J., Systems, M., May, A., Tumer, K., Yolum, P., Sonenberg, L., International, S. C. and Gerding, E. H. (2011), Online Mechanism Design for Electric Vehicle Charging, *in* 'In Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2011)', Taipei, Taiwan, pp. 811–818.

Feng, Y., Head, K. L., Khoshmagham, S. and Zamanipour, M. (2015), 'A real-time adaptive signal control in a connected vehicle environment', *Transportation Research Part C: Emerging Technologies* **55**, 460–473.

FHWA (2016), 'NGSIM Homepage'.
URL: *http://ngsim.fhwa.dot.gov*

Fleurbaey, M. (1994), 'On fair compensation', *Theory and Decision* **36**(3), 277–307.

Fleurbaey, M. (2008), To Envy or To Be Envied? Refinements of the Envy Test for the Compensation Problem, *in* 'Rational Choice and Social Welfare', Springer Berlin Heidelberg, Berlin, pp. 95–118.

Fleurbaey, M. and Maniquet, F. (2011), Chapter Twenty-Two – Compensation and Responsibility, *in* 'Handbook of Social Choice and Welfare', Vol. 2, pp. 507–604.

Florian, M. and Gaudry, M. (1980), 'A Conceptual Framework for the Supply Side in Transportation Systems', *Transportation Research Part B: Methodological* **14B**, 1–8.

Florian, M. and Gaudry, M. (1983), 'Transportation Systems Analysis: Illustrations and Extensions of a Conceptual Framework', *Transportation Research Part B: Methodological* **17B**(2), 147–153.

Friedman, E. J. and Parkes, D. C. (2003), 'Pricing WiFi at Starbucks: issues in online mechanism design', *Proceedings of the 4th ACM conference on Electronic commerce - EC '03* p. 240.

Furuhata, M., Daniel, K., Koenig, S., Ordonez, F., Dessouky, M., Brunet, M.-E., Cohen, L. and Wang, X. (2015), 'Online cost-sharing mechanism design for demand-responsive transport systems', *Intelligent Transportation Systems, IEEE Transactions on* **16**(2), 692–707.

Furuhata, M., Dessouky, M., Ordóñez, F., Brunet, M.-E., Wang, X. and Koenig, S. (2013), 'Ridesharing: The state-of-the-art and future directions', *Transportation Research Part B: Methodological* **57**, 28–46.

Garg, D., Narahari, Y. and Gujar, S. (2008*a*), 'Foundations of mechanism design: A tutorial Part 1-Key concepts and classical results', *Sadhana* **33**(2), 83.

Garg, D., Narahari, Y. and Gujar, S. (2008*b*), 'Foundations of mechanism design: A tutorial Part 2-Advanced concepts and results', *Sadhana* **33**(2), 131.

Gartner, N. H. (1983), *OPAC: A demand-responsive strategy for traffic signal control*, number 906.

Gaynor, T. (2015), 'Solving The Critical Mass Problem. What is The Simplest Thing That Could Work?'.

Ghoseiri, K. (2013), 'Dynamic Rideshare Optimized Matching Problem', *Ph.D. Dissertation at University of Maryland* .

Green, J. R. and Laffont, J.-J. (1979), *Incentives in Public Decision Making*, North-Holland, Amsterdam.

Groves, T. (1973), 'Incentives in Teams', *Econometrica* **41**(4), 617–631.

Guruswami, V., Hartline, J. D., Karlin, A. R., Kempe, D., Kenyon, C. and McSherry, F. (2005), 'On profit-maximizing envy-free pricing', *Proc. 16thSymp. on Discrete Algorithms (SODA)* pp. 1164–1173.

Gustafsson, A., Johnson, M. D. and Roos, I. (2005), 'The effects of customer satisfaction, relationship commitment dimensions, and triggers on customer retention', *Journal of marketing* **69**(4), 210–218.

Hafalir, I. E. (2007), 'Efficiency in coalition games with externalities', *Games and Economic Behavior* **61**(2), 242–258.

Hagerty, K. M. and Rogerson, W. P. (1987), 'Robust trading mechanisms', *Journal of Economic Theory* **42**(1), 94–107.

Hart, S. and Kurz, M. (1983), 'Endogenous Formation of Coalitions', *Econometrica* **51**(4), 1047.

Hartline, J. D. (2014), *Mechanism Design and Approximation*, unpublishe edn, Unpublished Book.

Hartline, J. D. and Yan, Q. (2011), 'Envy, Truth, and Profit', *EC'11: 12th Conference on Electronic Commerce - ACM SIGecom* .

Heinrich, S. (2010), 'Implementing real-time ridesharing in the San Francisco Bay Area', *Ph.D. Dissertation at San Jose State University* .

Henry, J.J., Farges, J.L., Tuffal, J. (1984), 'The PRODYN real time traffic algorithm', *Proceedings of the 4th IFAC-IFIP-IFORS conference on Control in Transportation Systems* (June), 307–311.

Herbawi, W. and Weber., M. (2011*a*), 'Ant colony vs. genetic multiobjective route planning in dynamic multi-hop ridesharing', *in Tools with Artificial Intelligence (ICTAI), 2011 IEEE Conference* .

Herbawi, W. and Weber., M. (2011*b*), 'Comparison of Multiobjective Evolutionary Algorithm for Solving The Multiobjective Route Planning in Dynamic Multi-hop Ridesharing', *Proceedings of the 11th European conference on Evolutionary computation in combinatorial optimization, Torino, Italy* pp. 84–95.

Herbawi, W. and Weber, M. (2011*c*), Evolutionary multiobjective route planning in dynamic multi-hop ridesharing, *in* 'European Conference on Evolutionary Computation in Combinatorial Optimization', Springer, pp. 84–95.

Herbawi, W. and Weber., M. (2012), 'A genetic and insertion heuristic algorithm for solving the dynamic ridematching problem with time windows', *In Proceedings of the fourteenth international conference on Genetic and evolutionary computation conference* .

Ichiishi, T. (1981), 'A Social Coalitional Equilibrium Existence Lemma', *Econometrica* **49**(2), 369.

Kamar, E. and Horvitz, E. (2009), Collaboration and Shared Plans in the Open World : Studies of Ridesharing, *in* 'Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence', pp. 187–194.

Kelly, K. L. (2007), 'Casual carpooling-enhanced', *Journal of Public Transportation* **10**(4), 6.

Kim, C. O., Park, Y. and Baek, J.-G. (2005), Optimal Signal Control Using Adaptive Dynamic Programming, *in* O. Gervasi, M. L. Gavrilova, V. Kumar, A. Laganá, H. P. Lee, Y. Mun, D. Taniar and C. J. K. Tan, eds, 'BT - Computational Science and Its Applications – ICCSA 2005: International Conference, Singapore, May 9-12, 2005, Proceedings, Part IV', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 148–160.

Kleiner, A., Nebel, B. and Ziparo, V. (2011), A Mechanism for Dynamic Ride Sharing based on Parallel Auctions, *in* 'Proceedings of the 22th International Joint Conference on Artificial Intelligence (IJCAI)', p. 7.

Kolm, S. (1995), 'Kolm, the economics of social sentiment, the case of envy', *The Japanese Economic Review* **46**, 63–87.

Krishna, V. and Perry, M. (1998), 'Efficient mechanism design', *Available at SSRN: http://dx.doi.org/10.2139/ssrn.64934* pp. 1–19.

Levin, M. W. and Boyles, S. (2015), 'Intersection auctions and reservation-based control in dynamic traffic assignment', *Transportation Research Record* **2497**(January), 35–44.

Li, J., Embry, P., Mattingly, S. P., Sadabadi, K. F., Rasmidatta, I. and Burris, M. W. (2008), 'Who Chooses to Carpool and Why?: Examination of Texas Carpoolers', *Transportation Research Record* **2021**, 110–117.

Lloret-Batlle, R. and Jayakrishnan, R. (2016), 'Envy-minimizing pareto efficient intersection control with brokered utility exchanges under user heterogeneity', *Transportation Research Part B: Methodological* **94**, 22–42.

Lloret-Batlle, R. and Jayakrishnan, R. (2017), Envy-free Pricing for Collaborative Consumption of Supply in Transportation Systems, *in* 'Transportation Research Procedia', Vol. 23, pp. 772–789.

Lloret-Batlle, R. and Jayakrishnan, R. (2018), Study of a Dynamic Cooperative Trading Queue Routing Control Scheme for Freeways and Facilities with Parallel Queues, *in* 'Submitted at Transportation Research Board 97th Annual Meeting. January 7-11, Washington, D.C.'.

Lloret-Batlle, R., Masoud, N. and Nam, D. (2017), 'Peer-to-Peer Ridesharing with Ride-Back on High-Occupancy-Vehicle Lanes: Toward a Practical Alternative Mode for Daily Commuting', *Transportation Research Record: Journal of the Transportation Research Board* (2668), 21–28.

Lowrie, P. (1990), *SCATS, Sydney co-ordinated adaptive traffic system: A traffic responsive method of controlling urban traffic*, Roads and Traffic Authority NSW, Darlinghurst, NSW Australia.

Mackie, P. J., Wardman, M., Fowkes, A. S., Whelan, G., Nellthorp, J. and Bates, J. (2003), 'Values of travel time savings UK', *Working Paper. Institute of Transport Studies, University of Leeds , Leeds, UK.* .

Mas-Colell, A., Whinston, M. D. and Green, J. R. (1995), *Microeconomic theory*, first edn, Oxford University Press.

Masoud, N. and Jayakrishnan, R. (2015*a*), A decomposition Algorithm to solve the multi-hop peer-to-peer ride-matching problem, *in* 'Transportation Research Board 94th Annual Meeting', number 15-2755.

Masoud, N. and Jayakrishnan, R. (2015*b*), A Real-Time Algorithm to Solve the Peer-to-Peer Ride-Matching Problem in a Flexible Ridesharing System, Technical report.

Masoud, N. and Jayakrishnan, R. (2017), 'A decomposition algorithm to solve the multi-hop Peer-to-Peer ride-matching problem', *Transportation Research Part B: Methodological* **99**, 1–29.

Masoud, N. and Lloret-Batlle, R. (2016), Increasing Ridership and User Permanence in Ridesharing Systems Using Novel Peer-to-Peer Exchange Mechanism, *in* 'Transportation Research Board 95th Annual Meeting', Washington DC.

Masoud, N., Lloret-Batlle, R. and Jayakrishnan, R. (2017), 'Using bilateral trading to increase ridership and user permanence in ridesharing systems', *Transportation Research Part E: Logistics and Transportation Review* **102**, 60–77.

McQuillin, B. (2009), 'The extended and generalized Shapley value: Simultaneous consideration of coalitional externalities and coalitional structure', *Journal of Economic Theory* **144**(2), 696–721.

Mirchandani, P. and Head, L. (2001), 'A real-time traffic signal control system: Architecture, algorithms, and analysis', *Transportation Research Part C: Emerging Technologies* **9**(6), 415–432.

Morency, C. (2007), 'The ambivalence of ridesharing', *Transportation* **34**(2), 239–253.

Myerson, R. B. and Satterthwaite, M. A. (1983), 'Efficient mechanisms for bilateral trading', *Journal of Economic Theory* **29**(2), 265–281.

National Research Council Transportation Research Board (2010), *HCM 2010: Highway Capacity Manual*.

Newell, G. (1989), *Theory of highway traffic signals*, University of California at Berkeley, Institute of Transportation Studies.

Newell, G. (1998), 'The rolling horizon scheme of traffic signal control', *Transportation Research Part A: Policy and Practice* **32**(1), 39–44.

Nisan, N. and Ronen, A. (2001), 'Algorithmic Mechanism Design', *Games and Economic Behavior* **35**(1-2), 166–196.

Nisan, N. and Ronen, A. (2007), 'Computationally feasible VCG mechanisms', *Journal of Artificial Intelligence Research* **29**, 19–47.

Nisan, N., Roughgarden, T., Tardos, E. and Vazirani, V. (2007), *Algorithmic Game Theory*, 1st edn, Cambridge University Press.

Park, J. Y. (2009), Network-wide Signal Control with Distributed Real-time Travel Data, PhD thesis, University of California, Irvine.

Parkes, D., Kalagnanam, J. and Eso, M. (2002), Achieving Budget-Balance with VCG-Based Payment Schemes in Combinatorial Exchanges, *in* 'In Proc. 17th International Joint Conference on Artificial Intelligence (IJCAI-01)', pp. 1161–1168.

Peleg, B. and Sudhölter, P. (2007), *Introduction to the theory of cooperative games*, Vol. 34, second edn, Springer Science & Business Media.

Pfeifer, P. E. (2005), 'The optimal ratio of acquisition and retention costs', *Journal of Targeting, Measurement and Analysis for Marketing* **13**(2), 179–188.

Quadstone Ltd (2010), 'Quadstone Paramics 6.7.2'.

R Development Core Team, R. (2014), 'R: A Language and Environment for Statistical Computing'.
URL: *http://www.r-project.org*

Ranaweera, C. and Prabhu, J. (2003), 'On the relative importance of customer satisfaction and trust as determinants of customer retention and positive word of mouth', *Journal of Targeting, Measurement and Analysis for marketing* **12**(1), 82–90.

Raney, S. (2010), 'San Francisco to Silicon Valley, California, Instant Ridesharing with Transfer Hub', *Transportation Research Record: Journal of the Transportation Research Board* (2143), 134–141.

Regue, R., Masoud, N. and Recker, W. (2016), Car2work: A Shared Mobility Concept to Connect Commuters with Workplaces, *in* 'Transportation Research Board 95th Annual Meeting', number 16-3184.

Robertson, D. and Bretherton, R. (1974), Optimum control of an intersection for any known sequence of vehicle arrivals, *in* 'IFAC/IFIP/IFORS Symposium on Traffic Control and Transportation Systems', Monte Carlo, Monaco, pp. 3–17.

Rust, R. T. and Zahorik, A. J. (1993), 'Customer satisfaction, customer retention, and market share', *Journal of retailing* **69**(2), 193–215.

Samra, S., El-Mahdy, A., Gomaa, W., Wada, Y. and Shoukry, A. (2011), 'Efficient parallel implementations of controlled optimization of traffic phases', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **7016 LNCS**(PART 1), 270–281.

Schepperle, H. and Böhm, K. (2007), 'Agent-Based Traffic Control Using Auctions', *Cooperative Information Agents XI* **LNCS 4676**, 119–133.

Schepperle, H. and Böhm, K. (2008), Auction-Based Traffic Management: Towards Effective Concurrent Utilization of Road Intersections, *in* '2008 10th IEEE Conference on E-Commerce Technology and the Fifth IEEE Conference on Enterprise Computing, E-Commerce and E-Services', IEEE, pp. 105–112.

Schrank, D., Eisele, B., Lomax, T. and Bak, J. (2015), 'Urban mobility scorecard', *College Station: Texas A&M Transportation Institute and INRIX* .

Sen, S. and Head, K. L. (1997), 'Controlled Optimization of Phases at an Intersection', *Transportation Science* **31**(1), 5–17.

Shapley, L. S. (1953), 'A value for n-person games', *Contributions to the Theory of Games* **2**(28), 307–317.

Shapley, L. S. and Shubik, M. (1966), 'Quasi-Cores in a Monetary Economy with Nonconvex Preferences', *Econometrica* **34**(4), 805.

Shelby, S. (2004), 'Single-Intersection Evaluation of Real-Time Adaptive Traffic Signal Control Algorithms', *Transportation Research Record* **1867**(1), 183–192.

Shelby, S. G. (2001), Design and Evaluation of Real-Time Adaptive Traffic Signal Control Algorithms, PhD thesis, University of Arizona.

Shen, W., Lopes, C. V. and Crandall, J. W. (2013), An Online Mechanism for Ridesharing in Autonomous Mobility-on-Demand Systems Masdar Institute of Science and Technology, *in* 'Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)', pp. 475–481.

Small, K. a. (1982), 'The scheduling of consumer activities: work trips.', *American Economic Review* **72**(3), 467–479.

Small, K. A. and Verhoef, E. T. (2007), *The economics of urban transportation*, second edn, Routledge.

Söderlund, M. (1998), 'Customer satisfaction and its consequences on customer behaviour revisited: The impact of different levels of satisfaction on word-of-mouth, feedback to the supplier and loyalty', *International Journal of Service Industry Management* **9**(2), 169–188.

Teodorović, D. and Dell'Orco, M. (2005), Bee colony optimization–a cooperative learning approach to complex transportation problems, *in* 'Advanced OR and AI Methods in Transportation: Proceedings of 16th Mini–EURO Conference and 10th Meeting of EWGT (13-16 September 2005)', Publishing House of the Polish Operational and System Research, Poznan, pp. 51–60.

The Wall Street Journal (2016), 'What's hot off the lots'.
URL: *http://online.wsj.com/mdc/public/page/2_3022-autosales.html*

Tiptopjob.com (2004), 'Survey: How Often Are You Late For Work?'.
URL: *http://www.tiptopjob.com/displaycontent/sectionid/27/contentid/384_-survey-how-often-are-you-late-for-work*

Varian, H. R. (2007), 'Position auctions', *International Journal of Industrial Organization* **25**(6), 1163–1178.

Vasirani, M. and Ossowski, S. (2008), 'Towards reservation-based intersection coordination: An economic approach', *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC* pp. 1–6.

Vasirani, M. and Ossowski, S. (2010), 'A Market-Based Approach to Accomodate User Preferences in Reservation-Based Traffic Management', *International Conference on Autonomous Agents and Multiagent Systems* pp. 103–110.

Vasirani, M. and Ossowski, S. (2012), 'A market-inspired approach for intersection management in urban road traffic networks', *Journal of Artificial Intelligence Research* **43**, 621–659.

Vickrey, W. (1961), 'Counterspeculation, Auctions, and Competitive Sealed Tenders', *Journal of Finance* **16**(1), 8–37.

Wang, X. (2013), Optimizing ride matches for dynamic ride-sharing systems, PhD thesis, Georgia Institute of Technology.

Wolfler Calvo, R., de Luigi, F., Haastrup, P. and Maniezzo, V. (2004), 'A distributed geographic information system for the daily car pooling problem', *Computers & Operations Research* **31**(13), 2263–2278.

Xu, S. X. and Huang, G. Q. (2014), 'Efficient auctions for distributed transportation procurement', *Transportation Research Part B: Methodological* **65**, 47–64.

Yang, I. and Jayakrishnan, R. (2015), 'Real-time network-wide traffic signal optimization considering long-term green ratios based on expected route flows', *Transportation Research Part C: Emerging Technologies* **60**, 241–257.

Yougov.us (2014), 'Nearly half of Americans are never late for work'.
URL: *https://today.yougov.com/news/2014/11/20/nearly-half-americans-are-never-late-work/*

Zhao, D., Zhang, D. and Gerding, E. H. (2014), 'Incentives in Ridesharing with Deficit Control', *13th International Conference on AAMAS 2014* pp. 1021–1028.

Appendix A

Appendix. Mathematical notation used in PEXIC

θ_i	value of delay savings (VDS) for agent i
t_i	departure time of agent i
a_i	arrival time of agent i
d_i	control delay of agent i
I_h	set of vehicles participating in rolling horizon (RH) h
$p_{i,h}$	price charged to agent i during RH h
$\pi_{i,h}$	total price charged to agent i after RH h
X_j^{min}, X_j^{max}	minimum and maximum barrier group length
$G_{min}^{p,r}$, $G_{max}^{p,r}$	minimum and maximum green length per phase p and ring r
$Y^{p,r}$	yellow and clearance time for phase p in ring r
s_j	state variable at stage j
$x_j(s_j)$	control variable at stage j given state variable s_j
$X_j(s_j)$	Set of control decision variables $x_j(s_j)$

$Z_j(s_{j-1})$	(alternative formulation) set of control decision variables $x_j(s_j)$
$z_{r,c}$	binary variable to select optimal phase order for ring r and phase combination c
$\phi_{n,c}$	binary variable to select optimal phase order for phase position n and phase combination c
$h_{sat}(i)$	Saturation headway for vehicle in ith position
v_j	value function at stage j
T	rolling horizon length
$L_j\left(s_j, x_j\right)$	vehicle list at stage j, after state s_j and control variable x_j
$D_j\left(s_j\right)$	departed vehicle list at stage j given state s_j
$Q_j\left(s_j\right)$	queued vehicle list at stage j given state s_j
$M_j(s_j)$	temporary optimal departure vehicle list at stage j given state s_j
J	optimal number of stages
V_h	optimal (final) vehicle list for RH h
$g_{r,c}$	green length on ring r and phase combination c
N_h	number of agents participating in RH h
ϵ_{ij}	envy term for agent i against agent j
γ_i	envy weight for agent i
$\overline{\theta}_h$, $\underline{\theta}_h$	maximum and minimum VDS within agents participating in RH h
a, b	logistic function parameters