COLLABORATION IN TRANSPORTATION

A Thesis
Presented to
The Academic Faculty

by

Okan Örsan Özener

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology
December 2008

COLLABORATION IN TRANSPORTATION

Approved by:

Dr. Özlem Ergun, Advisor
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Martin Savelsbergh, Advisor
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Alan Erera
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Mark Ferguson
School of Management
Georgia Institute of Technology

Dr. Pınar Keskinocak
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: 18 August 2008

To my family.

ACKNOWLEDGEMENTS

First, I would like to express my deepest gratitude to my advisors Dr. Özlem Ergun and Dr. Martin Savelsbergh, for their patience and guidance during this study. It has been a great pleasure and incredible experience to work under their supervision.

I would like to thank Dr. Alan Erera, Dr. Mark Ferguson, and Dr. Pınar Keskinocak for their time and effort in serving on my Ph.D. committee.

I thank to Ali Ekici, my officemate, for his patience and valuable comments throughout my study. Also, I thank my dear friends, Aykağan Ak, Serhan Duran, Murat Eröz, Ozan Gözbaşı, Burak Karacık, Fatma Kılınç Karzan, Elçin Kentel, Eser Kırkızlar, Gültekin Kuyzu, Melike Meterelliyoz Kuyzu, Evren Özkaya, Gökhan Samur, and Emrah Uyar for their friendship and support during my study.

Finally, I thank to my parents Güler and Tahsin Özener and my brother Ozan Önder Özener for their support and encouragement at every stage of my education.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

In this thesis, we investigate synergies between participants in transportation and distribution systems and we explore collaborative approaches to exploit these synergies to reduce transportation and distribution costs. Implementing collaborative approaches not only involves identifying and exploiting synergies between system participants, but also entails allocating any benefits fairly among them. We study collaboration in two environments: truckload transportation and vendor management inventory replenishment.

In truckload transportation synergies between loads may be exploited to increase truck utilization by reducing empty repositioning and thus reducing transportation costs. Shippers may exploit synergies by offering continuous move routes (with little or no empty repositioning) to carriers in return for reduced per mile charges. Carriers may exploit synergies by exchanging loads among themselves to reduce empty repositioning and thus increase truck utilization. In vendor managed inventory replenishment synergies between customers, due to their locations, usage rates, and storage capacities, may be exploited to reduce distribution costs by serving nearby customers on the same route at the same time.

By integrating non-cooperative and cooperative game theory methods with optimization techniques, we develop mechanisms to initiate and maintain collaborations. In contrast to earlier work on collaborative approaches, which focused on procurement, identifying and evaluating the value of synergies in transportation and distribution systems is difficult and requires optimization techniques. Furthermore,

generic game theory models seldom lead to practically viable collaboration mechanisms. Therefore, we take the practical characteristics of the environment into consideration when developing collaborative solutions, thus ensuring that the resulting mechanisms are viable in that environment.

The first part of the thesis addresses the cost allocation problem of a collaborative truckload transportation procurement network. More specifically, we study a logistics network where shippers identify collaborative routes with few empty truck movements in order to negotiate better rates with a common carrier. We investigate how to allocate the cost savings, or equivalently the cost, of these routes among the members of the collaboration. First, we show that proportional allocation methods and allocation methods used in practice have several drawbacks, particularly in terms of stability. Next, we develop several cost allocation mechanisms that satisfy desirable game theoretic and have desirable practical properties.

In the second part of the thesis, we investigate collaboration opportunities among carriers. When several carriers have to satisfy truckload transportation requests from various shippers, they may reduce their transportation costs by exchanging requests. First, we focus on computing the minimum cost to satisfy all requests, i.e., the cost associated with the "perfect" carrier collaboration. To be able to do so, we have to simultaneously assign requests to carriers and determine the optimal routes for satisfying the requests assigned to each carrier. Next, we develop and analyze various exchange mechanisms that allow carriers to exchange requests in order to realize some of the potential costs savings. As carriers act selfishly, it may not be possible to reach the perfect collaboration.

In the last part of the thesis, we study vendor managed inventory replenishment. Vendor managed inventory replenishment is a collaboration between a supplier and its customers. Our focus is on allocating the distribution costs incurred by the supplier among the customers. Simple cost allocation methods ignore synergies between the customers, due to their locations, usage rates, and storage capacities. As a result, the price charged to a customer for distribution does not represent the actual cost of serving that customer, and therefore the company may lose existing customers to the competition or decline prospective customers. We design a mechanism capable of computing a cost-to-serve for each customer that properly accounts for the synergies among customers.

CHAPTER I

INTRODUCTION

Collaboration is the process of working together to achieve a common objective that cannot be achieved by individual efforts. In a successful collaboration, parties with similar interests put forth a coordinated effort, which in turn will benefit all the parties participating in the collaboration. Usually, this involves agreeing on a common objective, devising a plan to realize that objective, and allocating responsibility to each collaborator. The degree of collaboration in a supply chain ranges from information sharing between the levels of the supply chain to strategic alliances between competitors.

The current trends in supply chain management, such as increased competition, need for improved customer responsiveness, outsourcing, and globalization drive companies towards collaboration. The companies realize the potential benefits of collaboration, including reduced costs, decreased lead times, and improved asset utilization and service levels, and are willing to collaborate with their suppliers, customers, and even competitors. In addition to these benefits, having a broader supply chain perspective enables firms to make better informed decisions on strategic issues.

Collaboration among companies occurs across a variety of levels and business functions. Collaborations in supply chains are categorized into horizontal and vertical collaborations. In horizontal collaborations, companies with similar characteristics, such as competitors, collaborate to achieve greater benefits. Group purchasing organizations are an example of horizontal collaboration where competitors collaborate to procure goods in order to receive volume discounts from a common seller. Capacity sharing among cargo carriers and seat capacity sharing among passenger airlines (code-sharing) are other examples of horizontal collaborations.

On the other hand, vertical collaboration refers to collaboration across different levels of the supply chain. Most of the collaborations in this category are based on information sharing and coordination of operations among buyers and sellers. A vast literature exists on supply chain contracts between buyers and sellers. These contracts are structured so as to accomplish supply chain coordination in order to improve the performance of the supply chain. A well-studied example of vertical collaboration is vendor managed inventory resupply where the seller is responsible for maintaining the buyer's inventory level in order to reduce inventory levels across the supply chain. These studies mainly focus on the inventory related costs rather than transportation costs.

Although collaborations offer potential benefits at every level, they also pose numerous challenges, which increase with the degree of collaboration. This is mostly because participants, though often working towards a common objective, are guided by their own self-interests. Therefore, any proposed mechanism to manage the collaboration's activities should yield collectively and individually desirable solutions. At the highest level, developing a successful collaboration involves two primary tasks: (1) identifying and exploiting synergies between individuals, and (2) allocating the resulting benefits among the collaborators.

Especially in transportation and distribution settings, where benefits depend on interactions between participants, identifying and exploiting synergies often involves solving complex optimization problems and thus in itself may be quite challenging.

On the other hand, cooperative game theory offers standard procedures to allocate the benefits of a collaboration. Cooperative game theory primarily focuses on allocating the benefits of a collaboration in a proper manner among the members. First, a collaboration should allocate the entire benefits achieved (or costs incurred) among the members. That is, a budget deficit or surplus is not desirable for a collaboration. If for some reason, e.g., a conflict with other restrictions placed on the benefit allocation, no allocation mechanism exists that achieves budget balance, approximate budget balance may be acceptable. Second, the collaboration should offer positive benefits to all members. That is, all participants must feel that it is in their best interest to enter a collaboration even with their competitors. Finally, to keep the collaboration together, an allocation mechanism should treat each participant "fairly." That is, each individual should be better off by being a member of the collaboration than by acting alone or by participating in another collaboration. Otherwise, a collaboration faces the risk of collapsing.

Besides these two primary tasks, several other issues affect the success of a collaboration, such as dealing with the dynamics of the business environment, satisfying industry imposed restrictions, and eliminating trust problems among participants.

In today's dynamic world of business, new companies may desire to enter a collaboration over time, thus increasing and changing the synergy between members. However, existing members may oppose the admission of the newcomers they would be worse off as a result of the expansion. This may lead to conflicts among existing members and a benefit allocation mechanism should aim to avoid or minimize these conflicts.

In addition to the general guidelines provided by cooperative game theory for managing a collaboration, practical or industry-imposed restrictions may play an important role in the performance of a collaboration as well. A benefit allocation mechanism may be perfect from a game theoretical perspective, but may possess undesirable or even unacceptable properties from a practical perspective. The ultimate goal in managing a collaboration is to develop a framework that simultaneously satisfies the game theoretic properties and the industry forced restrictions. However, in almost all situations satisfying all these properties and restrictions will not be possible. Therefore, the members of a collaboration should agree in advance on the set of desired properties and attempt to design a management strategy (e.g., a cost/benefit allocation method) that suits the given setting.

Another important issue in collaborative relationships is trust between selfish individuals. This is considerably more important in horizontal collaborations where competing individuals collaborate with each other. Due to this trust problem, information asymmetries exist between members. A substantial body of literature has examined information asymmetries among different agents in the supply chain including how to share information between those agents and how this affects the performance of the supply chain. Usually, information sharing benefits the collaborations and the collaborators, however this is not always the case.

1.1 Motivation

In this thesis, we discuss and analyze opportunities and challenges associated with collaborative approaches to managing transportation activities. We study collaboration in two environments: truckload transportation and vendor management inventory replenishment. A common property in these environments is that assessing synergies between the system participants is difficult and that allocating the benefits/costs is non-trivial.

The first application we study focuses on shippers' collaboration in truckload transportation. In this setting, synergies arise when shippers collaborate because of the increased geographic density of the set of shipments offered to a carrier. The carrier can reduce asset repositioning cost and can pass on some of the cost savings to the shippers. The second application we investigate is concerned with collaboration among carriers. Carriers may be able to increase the geographic density of the shipments they need to serve (and thus reduce their asset repositioning costs) by exchanging shipments with other carriers. The final application we analyze examines vendor managed inventory replenishment at a large industrial gas distributor. Distribution costs depend strongly on the synergies between customers in terms of their locations, usage rates, and storage capacities. With the right synergies fewer cost-effective routes are needed to serve the set of customers. For all three applications, the central focus is identifying and exploiting the synergies and developing mechanisms to allocate the benefits/costs to the system participants.

Being a \$623 billion industry, the trucking industry is a vital part of the U.S. economy. This annual revenue corresponds to 84.3% of the nations freight bill. The total volume of freight transported in 2005 was 10.7 billion tons accounting for 68.9% of the total freight volume in the United States [4]. According to the American Trucking Association, 26.2 million trucks are used for hauling this freight volume. The trucking industry is a labor-intensive industry. According to the Bureau of Labor Statistics, the trucking industry employed more than 1.6 million heavy-duty truck drivers and more than 0.94 million light-duty truck drivers in 2006.

The U.S. trucking industry is comprised of over 500,000 companies that provide transportation services for the shipment of goods. YRC Worldwide, Con-way, SIRVA Inc., J.B. Hunt Transportation Services and Schneider National Inc. are examples of large trucking companies (or carriers) operating in the United States [47]. The industry also contains numerous small carriers. The industry is highly fragmented in spite of the recent consolidation trends. As of 2005, 86% of the trucking companies operate six or fewer trucks, while 96% operate 28 or fewer trucks [4].

The trucking industry's operating expenses are higher than ever. Fuel prices, which accounts for a fourth of the industry's operating expenses, and highway road fees, two important contributors to operating expenses, are at an all time high. "Each penny increase in diesel costs the trucking industry 381 million over a full year" [3]. The trucking industry has a low profit margin mainly due to high operating costs.

The highly fragmented industry structure, intense competition, and low profit margins have forced many small trucking companies out of business. It has also fuelled the large number of mergers and acquisitions in the trucking industry. Large trucking companies are better able to handle the increased operational costs because of economies of scale and economies of scope.

Mergers and acquisitions increase the density of shipment requests across the carriers' networks, which results in high utilization of assets and, hence, low per mile operating cost for the carriers. A less drastic (and sometimes preferable) way to increase shipment density and asset utilization is collaboration. Several trucking companies already sought out collaboration to improve their overall performance. For example, six of the largest truckload carriers in the United States, namely J.B. Hunt Transport Services, Werner Enterprises, Swift Transportation, M.S. Carriers, U.S. Xpress Enterprises and Covenant Transport, agreed on the partnership under a collaborative logistics network called Transplace.com. This partnership allows the trucking companies and companies requesting transportation services to meet and match their hauling capacity and transportation demand using an internet based management technology [46]. Shippers also seek out collaborative solutions to reduce their transportation costs. For example, through another collaborative logistics network, Sterling TMS, two member shippers have identified a 2,500-mile continuous move route that resulted in a 19 percent savings for both shippers over their individual company based solution.

In summary, collaboration in truckload transportation can lead to significant cost saving opportunities. In the first and second part of the thesis, we explore ways to identify collaborative opportunities among shippers and among carriers and develop frameworks to successfully manage these collaborations.

In the first part of the thesis, we study a logistics network where shippers collaborate and bundle their shipment requests in order to negotiate better rates with a common carrier. In this setting, shippers identify collaborative continuous move cycles with decreased overall empty truck movements. In the absence of side constraints, determining the set of continuous move cycles minimizing the total cost of covering all the shippers' transportation requests (i.e., identifying and exploiting the synergies among the shippers) can be done in polynomial time. Benefit allocation in this setting is equivalent to allocating the total cost of the continuous move cycles among the members of the collaboration. We propose and analyze various mechanisms.

Logistics exchanges or collaborative logistics networks provide a common platform not only for shippers, but also for carriers to explore collaborative opportunities. In the second part of the thesis, we study a logistics network with multiple carriers offering transportation services to shippers (or to a shippers' collaboration). In this setting, carriers identify transportation commitments that when exchanged among them decrease overall empty truck movements and thus overall cost. The potential for cost savings is due to the differences in cost structures of the carriers. The differences in cost structures of the carriers is the result of various factors such as the size of the company, whether the company is a regional carrier or not, and any pre-existing contracts the company (which we will often refer to as the existing network). We focus first on identifying and fully exploiting the synergies in this multi-carrier setting, i.e., determining the system optimal solution from a centralized perspective. Then, we take the carriers' perspective and investigate how well exchanges of transportation commitments between carriers can approximate the system optimal solution. Note that almost always there will be synergies between transportation commitments across different carriers' networks. The underlying reason for this is that the carriers' valuation of a transportation commitment changes with every new transportation commitment in the system or with every change of the cost structure of a carrier. As a result, unilateral or bilateral exchanges may lead to a win-win situation for the carriers. We develop mechanisms to help the carriers exchange transportation commitments and to realize the benefits of the collaboration.

In the last part of the thesis, we study a separate but related distribution setting: vendor managed inventory resupply. In vendor managed inventory resupply, a supplier and its customers collaborate to reduce inventory holding and distribution costs. Collaboration occurs because customers transfer control of their inventory management to the supplier. Our research is motivated by our relationship with a large industrial gas company that operates a vendor managed inventory resupply policy. The company replenishes the storage tanks at customer locations by a homogenous fleet of tanker trucks under the company's control. The customers and trucks are assigned to a specific facility, and tanker trucks and storage tanks contain only a particular type of product. Our goal in this setting is to allocate the distribution costs among the customers (the company incurs the inventory holding costs at the production facility and the customer sites, so we can ignore inventory holding costs). That is, we want to compute a "cost-to-serve" for each customer. Cost-to-serve information is of value to the company for marketing and sales purposes or distribution planning purposes. If each customer is served individually, then determining the cost-to-serve is easy. When multiple customers are visited together on delivery routes, the situation becomes much more complex. The company has a simple cost allocation mechanism in place, but it ignores the synergies that may exist between customers, such as their locations, their usage rates, and their storage capacities. Hence, their cost-to-serve estimate is not very accurate. We develop and analyze various methods for computing a cost-to-serve that does take synergies among customers into account.

1.2 Contributions of the Thesis

In this thesis, we investigate synergies between participants in transportation and distribution systems and we explore collaborative approaches to exploit these synergies to reduce transportation and distribution costs. Implementing collaborative approaches not only involves identifying and exploiting synergies between system participants, but also entails allocating any benefits fairly among them. We study collaboration in two environments: truckload transportation and vendor management inventory replenishment.

The major contribution of this research is the integration of optimization and non-cooperative and cooperative game theory to address real-life collaborative logistics problems. A common characteristic of these problems is that assessing the synergies that drive the collaboration is a complex optimization problem. Furthermore, contrary to most studies on collaborations in the literature, we take into consideration the practical aspects of the application and the industry imposed restrictions as well as desirable properties from non-cooperative and cooperative game theory. In summary, we design implementable mechanisms with low computational times rather than well-studied generic and possibly exponential-time mechanisms as often proposed in the literature.

We are the first to study the benefit/cost allocation problem associated with a shippers' collaboration network and to devise cost allocation mechanisms that ensure the sustainability of the collaboration. The natural cost allocation mechanism to consider, and the one used in practice, is a proportional cost allocation method (e.g. proportional to individual costs). We show that proportional allocation methods have several drawbacks, particularly in terms of stability. That is, proportional cost allocation methods may allocate a subgroup of participants more than the subgroup's individual cost. The reason for this is that a proportional cost allocation ignores the synergies between transportation needs of the shippers. The computational experiments on real-world instances reveal that these excess charges correspond to 25% of the subgroup's individual cost on average. Therefore, using such a cost allocation method may result in a break-up of the collaboration.

We develop a cost allocation mechanism that is budget balanced and stable. Furthermore, unlike most studies in the literature on collaborative games, we define a new set of properties that are desirable from a practical perspective. We show that no cost allocation method can satisfy all the desired properties, and hence, devise methods that result in cost allocation methods that are best suited for a particular application. Our proposed methods perform significantly better than the proportional cost allocation methods. For instance, even our stability-relaxed cost allocation mechanism has an average excess charge of around 8%, approximately one third of that of the proportional cost allocation methods. We also show that our solution methodologies work under different carrier's pricing schemes.

We extend the Lane Covering Problem (LCP), the optimization problem associated with the shippers' collaboration problem, to the Multi-Carrier Lane Covering Problem (MCLCP), which allows us to study carriers' collaboration. Identifying and exploiting synergies in a multi-carrier setting involves simultaneously assigning requests to carriers and determining the optimal routes for satisfying the requests assigned to each carrier. Not surprisingly, the problem is NP-hard in all but a few special cases. We have designed and implemented several heuristic approaches and computational experiments show that large problem instances can be solved with acceptable quality and computation times.

We also analyze carrier collaboration opportunities in this setting. Carriers may be able to increase their profits (or decrease their costs) by exchanging lanes among them. We develop lane-exchange mechanisms based on Nash equilibrium concepts. The proposed mechanisms help carriers to decide which lanes to offer for exchange, how to price the lanes that are offered for exchange, and whether to accept lanes offered by other carriers. The exchange mechanisms differ depending on the level of information sharing between the carriers. The challenge in defining an effective lane-exchange mechanism is that it has to be acceptable to the self-goal seeking carriers, it has to be beneficial from a centralized perspective, and it cannot be computationally prohibitive. Especially without information sharing among the carriers, selecting a lane to offer and determining an associated side payment is non-trivial. With information sharing among carriers, the number of alternative strategies for the carriers is quite large, which makes analyzing the exchange process fairly complicated.

Although carriers' collaboration is beneficial from a centralized perspective, we present examples where at least one of the participants is always worse off. We also show a counterintuitive result: increasing the level of information sharing is not always beneficial for the collaboration. Last but not least, we show empirically that an exchange mechanism with information sharing and price determination yields satisfying results and almost achieves the centralized system optimal solution and that this can be done in an acceptable amount of computation time.

In the final part of the thesis, we consider the problem of computing the cost-to-serve for customers that are supplied under a vendor managed inventory policy. Cost allocation in routing and distribution problems has been rarely addressed in the literature. Cost allocation methods studied in the literature for related problems, such as the traveling salesman problem and the vehicle routing problem, do not have to deal with the time or inventory component of the problem cannot be applied directly. Furthermore, the methods proposed for the associated routing games are usually generic, such as the Nucleolus and the Shapley Value, and thus have exponential computation times and ignore any practical aspects of the problem. We show that the instability of such cost allocations can be arbitrary large. To the best of our knowledge, our work is the first to consider cost allocation of an inventory routing game. We integrate cooperative game theory and optimization methods to develop cost allocation methods that have low computational times and that satisfy the game theoretical aspects as well as the practical restrictions of the application.

CHAPTER II

ALLOCATING COSTS IN A COLLABORATIVE TRANSPORTATION PROCUREMENT NETWORK

2.1 Introduction

Because of increased pressure to operate more efficiently and satisfy ever-increasing demand from customers for better service, companies are forced to realize that suppliers, consumers, and even competitors can be potential collaborative partners. Thus, many companies use collaborative practices such as group purchasing and capacity and information sharing to reduce systemwide inefficiencies and cut operational costs. Recent developments in communications technology enable companies to consider a range of opportunities that becomes possible by collaborating with others.

In a traditional truckload logistics market, the shipper submits its freight requests to several carriers and negotiates terms with them. These requests consist of multiple *lanes* to be serviced. A *lane* corresponds to a shipment delivery from an origin to a destination with a full truckload. In truckload transportation, the truck moves directly from the origin to the destination without visiting any intermediate locations.

In contrast, a carrier collects the freight requests from several shippers and offers prices based on its existing lane network and the lanes it anticipates getting by the time of service. The shipper procures the transportation service from the carrier that offers the lowest price for the freight request. A key aspect that affects the carrier's operational cost is asset repositioning. Asset repositioning, equivalently deadheading, is empty truck movement from a delivery location to a pickup location. Carriers often reposition their assets to satisfy the demands of different shippers. Asset repositioning decreases the capacity utilization of the carrier, which increases operational costs.

According to the estimates of American Trucking Association, the ratio of empty mileage to total mileage for large truckload carriers is approximately 17% whereas the same ratio is approximately 22% for small truckload carriers; together this corresponds to approximately 35 million empty miles monthly [4]. Because these deadhead miles generally result from the imbalance of freight requests of different shippers, asset repositioning costs are reflected in the lane prices.

A collaboration for procurement of transportation services is established when shippers get together to minimize their total transportation costs by better utilizing the truck capacity of a carrier. Nistevo, Transplace, and One Network Enterprises are examples of collaborative logistics networks that enable shippers and carriers to manage their transportation activities. A variety of companies such as General Mills, Georgia-Pacific, and Land O'Lakes can identify routes with less asset repositioning for their transportation needs, which result in considerable savings in their transportation expenses by being members of such collaborative logistics networks. For example, after forming collaborative partnerships with others in the Nistevo Network, Georgia-Pacific's percentage of empty movements decreased from 18% to 3%, which corresponds to $11,250,000 savings yearly [42].

The collaborative setting we consider in this chapter is relevant for companies that regularly send truckload shipments, say several days of the week, and are looking for collaborative partners in similar situations to cross utilize a dedicated fleet. The truckload shipments that participate in such a collaboration are executed periodically. Since dedicated fleets are used for such shipments and the trucks are expected to return back to the initial location at the end of each period, the shippers are responsible for the anticipated asset repositioning cost. In practice, different carriers may have different costing schemes associated with empty truck movements. For instance, the anticipated asset repositioning cost along a lane can be a percentage of the lane cost and this percentage may be different for each lane of the network. The carrier may also charge a fixed amount for asset repositioning. Initially, we assume that the carrier charges a specific fraction of the original lane cost for each asset repositioning. We call this fraction the "*deadhead coefficient*" (θ). We will show in Section 2.6 that the proposed cost allocation procedures are valid for other commonly used pricing structures as well.

The total asset repositioning costs incurred by a carrier depend on the entire set of lanes served by the carrier at the time of service. However, when a carrier gives a price to a shipper lane neither the carrier nor the shipper has perfect information on the final lane network the carrier will cover at the time of service. Hence, the price the carrier offers to a shipper includes a mark-up due to the expected repositioning cost associated with the shipper's lane. On the other hand, if the shipper is able to bundle its lanes with complementary ones and provide a continuous move with minimal repositioning at the time of purchase, then the shipper will be able to negotiate for better rates from the carrier.

In this study, we first show that the set of routes that cover all the lanes of the shippers' collaboration with minimum cost can be identified by solving a linear program. Subsequently, we investigate the question of allocating the total cost of these routes among the members of the collaboration. Although the benefits of forming collaborations are appealing, ensuring the sustainability of a collaboration is the key for realizing these benefits. A successful membership mechanism should distribute the joint benefits/costs of collaborating among the members of the collaboration in a "stable manner" (the concept of stability will be defined in details in Section 2.3). Current practice in truckload transportation markets is to allocate joint costs proportionally to the cost of servicing the participating shipper's lanes before collaboration. We show that although such proportional allocation schemes are easy to implement, they are not stable from a game theoretic point of view. Furthermore, we demonstrate that the potential benefits from a shippers' collaboration depend on the synergies among the lanes, hence, even a small modification in the lane network may cause the cost allocated to each shipper to change significantly. We then attempt to design cost allocation mechanisms which are stable, encourage the expansion of the collaboration, and guarantee each shipper to pay no more than the cost allocated before the expansion.

"In the following, we will use the terms "*original lane cost*" to designate the cost of performing a full truckload delivery from a lane's origin to its destination and "*stand alone cost of a set of lanes*" to designate the minimum cost of serving these lanes (the original lane costs plus the asset repositioning costs) independent of the lanes that are not in this set. When the set of lanes corresponds to the set of lanes requested by a shipper, the corresponding stand alone cost is referred to as the "*stand alone cost of the shipper*". Note that the stand alone cost of a shipper can be less than the sum of the stand alone costs of its lanes due to the synergy between these lanes. Finally the "*cost of the shippers' collaboration*" or equivalently the "*total cost of the collaboration*" or the "*total budget*" is the total cost of servicing all shippers' lanes in the collaboration.

Next, we demonstrate the challenges associated with the shippers' collaboration cost allocation problem with an example. Although the example presented in Figure 1 is fairly simple, it nevertheless shows that the basic cost allocations may not be desirable from both practical and game theoretic points of view and finding a desirable cost allocation method is non-trivial.

Consider a shippers' network (Figure 1) with one shipper (A) with a lane from node 1 to node 2 and another shipper (B) with a lane from node 2 to node 1. Suppose that the cost of covering a lane with a full truckload is equal to 1. We assume the deadhead coefficient, θ, to be 1, then the cost of covering a lane with an empty truck is equal to 1 as well. Then the total cost of covering the lanes in this network is 2. The stand alone cost of each lane in this example is equal to 2. Since each shipper

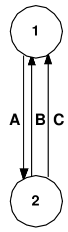

Figure 1: A shippers' network.

has only one lane in this example stand alone cost of each shipper is equal to 2 as well.

In a proportional cost allocation method, the costs are allocated to the lanes (or to the shippers) proportional to the stand alone costs of the lanes (or shippers). In our example, each shipper has only one lane, therefore both methods yield the same result and allocate a cost of 1 to each shipper.

If a new shipper (C) with a lane from node 2 to node 1 enters the collaboration, the total cost of covering the lanes in the network becomes 4. All three lanes (shippers) have the same stand alone cost. Then, the proportional cost allocation method allocates a cost of $\frac{4}{3}$ to each lane (shipper). However, with this allocation, it is not hard to see that shippers A and B (equivalently A and C) are better off collaborating on their own with a total cost of 2. Therefore, the proportional cost allocation in this case is not stable and a subgroup, namely A and B (or A and C), has an incentive to part company with C (or B) and cooperate on its own.

As we consider all possible cost allocations, we conclude that the only allocation where the grand coalition is not threatened by any subgroup of its members is the allocation of $(0, 2, 2)$ to shippers A, B and C, respectively. Since shipper A has a higher bargaining power compared to the other two shippers, it is expected to be charged less than B and C even though they have the same stand alone costs. Moreover, under any cost allocation method where shipper A has a positive allocation, either one of the other shippers will be willing to cover some of the expenses of shipper A in order to get into a coalition with A. However, charging shipper A nothing makes A a *free-rider* which may not be desirable in a collaboration. Furthermore, in the only stable allocation, where the grand coalition is maintained together, both shippers B and C are allocated their stand alone costs, so being in a collaboration brings no positive value for these two shippers.

Thus, we conclude that even on a very simple example, basic cost allocation methods such as proportional or stable cost allocation might have some undesirable properties for the shipper collaboration problem and designing a cost allocation method that has all the desired properties may not be possible. Accordingly, one must choose a set of desirable properties and design allocation methods that suit the problem at hand and this is the motivation behind our work.

The rest of this chapter is organized as follows. In Section 2.2, we review the related literature. In Section 2.3, we briefly discuss well studied cost allocation properties from cooperative game theory. In Section 2.4, we first give a formal statement of the shippers' collaboration problem and then develop a stable cost allocation method that allocates the total cost (i.e. an allocation in the *core*) for the shippers' collaboration problem using linear programming duality and discuss three other allocation schemes: the *nucleolus*, the *Shapley Value* and *cross monotonic* cost allocations. In Section 2.5, we develop cost allocations, by relaxing the budget balanced and stability properties, that do not allow the costs allocated to be less than the original lane costs and guarantee a reduction from the stand alone cost to each shipper. In Section 2.6, we investigate how different carrier pricing schemes affect our solution methodologies. In Section 2.7, we computationally demonstrate how our methods perform on several classes of randomly generated and real-life instances. Concluding remarks are provided in Section 2.8.

2.2 Literature Review

There are two streams of literature in cooperative game theory and truckload transportation that are relevant to our work. Cooperative game theory studies the class of games in which selfish players form coalitions to obtain greater benefits. A major portion of the literature on cooperative games focuses on finding cost allocations that are stable and allocate the total cost to the members.

We refer the reader to Young [48] that gives a thorough review of basic cost allocation methods and to Borm et al. [7] for a survey of cooperative games associated with operations research problems. The games generated by linear programming optimization problems, which are relevant to the game we consider in this chapter, are studied in Owen [31]. Owen [31] proves that for such games a cost allocation that is stable and allocates the total cost can be computed from an optimal solution to the dual of the linear program. Owen [31] also shows that for the linear production game it considered the converse does not hold and gives an example where such a cost allocation is not included in the optimality set of the dual linear program. These results are further extended in Kalai and Zemel [25], Samet and Zemel [35], and Engelbrecht-Wiggans and Granot [14]. Kalai and Zemel [25] establish the correspondence of every such cost allocation to an optimal dual solution for a flow game over a simple network. Samet and Zemel [35] and Engelbrecht-Wiggans and Granot [14] generalize this result and extend it to some LP games that include the games in which this correspondence is known to exist.

The contribution of this study is to provide a general framework for the cost allocation mechanisms that can be used to manage real-life collaborative truckload transportation networks. There exists a set of papers that join cooperative game theory and classic routing problems, such as Engevall et al. [16], which consider a vehicle routing game, Sanchez-Soriano et al. [36], which study transportation games where buyers and sellers are disjoint sets, Granot et al. [22], which analyze delivery games that are associated with the Chinese postman problem, Derks and Kuipers [12], which study routing games, and Tamir [44], which considers continuous and discrete network synthesis games. These papers in general study the existence of cost allocations with well-studied properties from cooperative game theory and propose computational procedures for finding such allocations. However, to the best of our knowledge there is no literature on cost allocation mechanisms for collaborative logistics problems that identify what the relevant cost allocation properties for the given application should be and develop mechanisms based on these properties. In this chapter while searching for a cost allocation method, we do not restrict ourselves only to the concepts from cooperative game theory but also consider the requirements relevant for collaborative truckload transportation networks.

On the other hand, the problem of finding efficient routes, continuous paths and tours, that minimize asset repositioning costs in a collaborative truckload transportation network has been studied by Moore et al. [28], Ergun et al. [18], and Ergun et al. [17]. These papers study the underlying optimization problems with side constraints such as temporal and driver restrictions and propose heuristic algorithms for solving them.

2.3 Preliminaries from Cooperative Game Theory

In this section, we discuss some of the well-known cost allocation properties from the cooperative game theory literature.

Cooperative game theory analyzes situations where a group of players act together in order to obtain certain benefits. When studying benefit/cost allocation related to cooperation, there is often a need for transferring benefits/costs from one player to another. Under the existence of a common transferable commodity such as money, the players may share benefits/costs of the collaboration among themselves in any possible way (if there are no further restrictions). A collaborative game where the players are allowed to compensate each others' costs with side payments, is called a *transferable payoffs* game.

In a *budget balanced*, or *efficient*, cost allocation, the total cost allocated to the members of the collaboration is equal to the total cost incurred by the collaboration. That is, a budget deficit or a surplus is not created. Most allocation methods studied in the literature attempt to find budget balanced allocations. However, in some games, budget balance property conflicts with a more desirable property. In such games, it is possible to seek *approximate budget balanced* cost allocations that recover at least α-percent of the cost (See Pal and Tardos [32] and Jain and Vazirani [24] for two applications).

In a *stable* cost allocation, no coalition of members can find a better way of collaborating on their own. Hence the grand coalition is perceived as stable and is not threatened by its sub-coalitions. Thus, stability is the key concept that holds a collaboration together. The set of cost allocations that are budget balanced and stable is called the *core* of a collaborative game. For a collaborative game, the *core* may be empty, that is, a budget balanced and stable cost allocation may not exist. Although stability is a key aspect in establishing a sustainable collaboration, it is not altogether meaningless to consider cost allocation methods with relaxed stability constraints. Due to the costs associated with managing collaborations, limited rationality of the players and membership fees, a sub-coalition might not be formed even though it offers additional benefits to its members. Also, for a collaborative game with an empty core, either budget balance or stability condition should be relaxed in order to find a cost allocation. Therefore, relaxing the stability restriction in a limited way might be acceptable for a cost allocation method.

When there are multiple cost allocations in the core, some of these allocations might be more desirable than the others. One such allocation that is well-studied in the cooperative game theory literature is the *nucleolus*. Nucleolus, introduced by

Schmeidler [37], is the cost allocation that lexicographically maximizes the minimal gain, the difference between the stand alone cost of a subset and the total cost allocated to that subset, over all the subsets of the collaboration. If the core is nonempty, the nucleolus is included in the core. Nucleolus may still exist even though the core is empty if there exists a cost allocation that is budget balanced and allocates costs to the players that are less than or equal to their stand alone costs.

A well-known cost allocation method is the *Shapley Value*, which is defined for each player as the weighted average of the player's marginal contribution to each subset of the collaboration [38]. Shapley Value can be interpreted as the average marginal contribution each member would make to the grand coalition if it were to form one member at a time [48]. Even if the core is non-empty, Shapley Value may not be included in the core.

After the collaboration is initiated, over time there may be potential new members willing to enter the collaboration. Although, it is expected that expansion will be beneficial for the collaboration as a whole due to increased synergies between the members, these benefits may not be distributed evenly among the members. In fact, some of the existing members of the collaboration may be worse off due to the addition of new members under some cost allocation methods. Consequently, these members may oppose the admission of the newcomers which, in turn, forfeits the additional benefits to the collaboration. To avoid this, it is preferable to have a *cross monotonic* cost allocation method in which any member's benefit does not decrease with the addition of a newcomer. A cross monotonic cost allocation is also stable if the allocation is budget balanced [29].

Having a cross monotonic cost allocation is favorable for contractual agreements. If a cross monotonic cost allocation method is used for distributing the costs, then at the contracting stage, a specific cost can be allocated to each member of the collaboration and the members will be assured of never paying more than that value as long as no one leaves the collaboration. See Tazari [45] for a survey of cross monotonic cost allocation schemes. Unfortunately, as will be discussed, the cross monotonicity property turns out to be a very restrictive one in our setting.

The *equal treatment of equals property* ensures that two participants of a collaboration have the same allocation if they are identical in every aspect relevant to the allocation problem in consideration. In the shippers' collaboration problem, equal treatment of equals does not imply that the costs allocated to the lanes that have the same stand alone cost should be the same. It ensures that the cost allocated to two different shippers is the same if they have lanes with same origin and destination. Throughout the chapter, this property is kept as an invariant.

A player i is a *dummy player* if the incremental cost of adding i to any coalition is zero and the *dummy axiom* states that a dummy player should get a cost allocation equal to zero.

A cost allocation method is *additive*, if for any joint cost functions, allocation of their sum is equal to the sum of their individual allocations.

Finally, there are some additional cost allocation properties we desire in our setting. As in the example given in the Introduction, under some cost allocation methods it is possible to have members, which pay less than their original lane costs or pay their stand alone cost. We believe that both of these situations might be considered undesirable. We study allocations with "minimum liability" restriction where the shippers are responsible for at least their original lane cost. We also study allocation methods where each shipper is guaranteed an allocation less than its stand alone cost so that being a member of the collaboration offers a "positive benefit" compensating the difficulties (such as integration) of collaborating. Note that, when either of these two restrictions is imposed, it is not possible to have a budget balanced and stable cost allocation for the shippers' collaboration problem as demonstrated in the example of the Introduction section. Hence, we relax the budget balance and stability properties in a limited way and develop allocations with the above two restrictions.

2.4 Cost Allocation Methods with Well-known Properties

In this section, we define the "*Lane Covering Problem*" that seeks to identify the optimal set of cycles covering the lanes in the collaboration and the cooperative game associated with this optimization problem. We also list our assumptions in this section. We then determine a cost allocation mechanism in the core of the shippers' collaboration problem using linear programming duality. Next, we briefly mention the nucleolus and an alternative cost allocation scheme, the Shapley Value. Finally, we discuss the relationship between the core of the game and cross monotonic cost allocations and devise allocations with this property.

2.4.1 Problem Definition

In the shippers' collaboration problem, we consider a collection of shippers each with a set of lanes that needs to be served. Given the cost of covering each lane and the anticipated repositioning charges, the collaboration's goal is to minimize the total cost of transportation such that the demand of each shipper in the collaboration is satisfied. The asset repositioning costs depend on the complete lane set of the shippers. As complementarity (synergy) of lanes from different shippers increases (i.e. they form continuous tours with no or minimal deadheading) gains from collaboration and the incentive to collaborate increase.

Next, we summarize our assumptions. First, we assume that every shipper is accepted to be a member of the collaboration even if a shipper does not create a positive value for any other shipper in the collaboration. The question of which shippers should be in the collaboration is beyond the scope of this work. Second, each lane corresponds to a truckload delivery between the origin and the destination of the lane. We assume that the cost of collaboration among the members is negligible. Also, a shipper is not required to submit all of its lanes to the collaboration. Hence, a shipper may exclude a subset of its lanes from the collaboration, and individually negotiate prices with the carrier for those lanes or form another collaboration with other shippers, if it is profitable to do so. We also assume that the lane set consists of only repeatable lanes which means that each shipper has the same lanes to be traversed each period. Finally, we assume that there are no side constraints (like time windows, driver restrictions ... etc.) on the problem.

We will refer to this total transportation costs minimization problem as the *Lane Covering Problem* (LCP). The LCP is defined on a complete directed Euclidian graph $G = (N, A)$ where N is the set of nodes $\{1, \ldots, n\}$, A is the set of arcs, and $L \subseteq A$ is the lane set requiring service. Let c_{ij} be the cost of covering arc (i, j) with a full truckload. The deadhead coefficient is denoted by θ, where $0 < \theta \leq 1$ hence the asset repositioning cost along an arc $(i, j) \in A$ is equal to θc_{ij}. Then the LCP, the problem of finding the optimal lane cover, can be solved by finding a solution to the following integer linear program:

$$P: \quad z_L(r) = \min \sum_{(i,j)\in L} c_{ij}x_{ij} + \theta \sum_{(i,j)\in A} c_{ij}z_{ij} \tag{1}$$

$$s.t. \sum_{j\in N} x_{ij} - \sum_{j\in N} x_{ji} + \sum_{j\in N} z_{ij} - \sum_{j\in N} z_{ji} = 0 \quad \forall i \in N \tag{2}$$

$$x_{ij} \geq r_{ij} \quad \forall (i,j) \in L \tag{3}$$

$$z_{ij} \geq 0 \quad \forall (i,j) \in A \tag{4}$$

$$x_{ij}, z_{ij} \in \mathbb{Z}. \tag{5}$$

In P, x_{ij} represents the number of times lane $(i, j) \in L$ is covered with a full truckload and z_{ij} represents the number of times arc $(i, j) \in A$ is traversed as a deadhead arc. We let r_{ij} be the number of times lane $(i, j) \in L$ is required to be covered with a truckload and r be the matrix of r_{ij}'s. Constraints (2) are flow balance constraints for all the nodes in the network. Constraints (3) ensure that the transportation requirement of each lane in L is satisfied.

Since P is a minimization problem, it is clear that variables x_{ij} will be equal to r_{ij} in any optimal solution and they can be deleted from P resulting in a simpler formulation. However, for the clarity of the following discussions, we work with the above formulation.

In this chapter, we focus on the cost allocation game generated by the LCP. In the cooperative game we consider, we designate the lanes in L as the players and try to allocate the total cost of the collaboration among all the lanes served. Since we assume that a shipper may decide to take out a subset of its lanes from the collaboration, allocating the cost to the lanes rather than to the shippers is more appropriate in our setting. The total cost allocated to a shipper is then the sum of the costs allocated to the lanes that belong to the shipper. The characteristic function $z_S^*(r^S)$ (or z_S^*) is the optimal cost of covering all the lanes in $S \subseteq L$.

For a given set of lanes, it can be very hard to identify the dynamics of an instance and contribution of each lane to the network without solving an LCP, since there are many subsets of lanes and determining the relationship between these subsets is a formidable challenge. Even with a small modification in the network, optimal set of cycles can change entirely requiring us to solve a new LCP to determine the new optimal set. Therefore, the cost allocations may change significantly with a small change in the input.

2.4.2 The Core of the Shippers' Collaboration Problem

As stated before, Owen [31] proves that the optimal dual solutions lead to core cost allocations for LP-games, cooperative games arising from a linear program. Kalai and Zemel [25] show that the same result holds for flow games, which can be transformed into the game discussed by Owen [31]. In the view of these results we first prove that LCP is a linear optimization problem, hence showing that the optimal dual solutions of LCP yield cost allocations in the core for the shipper collaboration problem. Next, we show that the converse of the statement also holds; every cost allocation in the core corresponds to an optimal dual solution unlike the game considered in Owen [31].

Lemma 1 *The core of the shippers' collaboration problem with transferable payoffs is non-empty. A cost allocation in the core can be constructed in polynomial time by solving D, the dual of the linear programming relaxation of P.*

Proof: P is a minimum cost circulation problem, hence solving its linear relaxation is sufficient to find an integer solution (see Ahuja et al. [1]). The structure of P is equivalent to the flow problem discussed by Kalai and Zemel [25], since there exist an arbitrary cost vector (c), service requirement constraints corresponding to each player in the collaboration (constraints (3)), and flow balance constraints for each node in the network (constraints (2)). Therefore the game we consider is an LP game. Hence, the core for this game is nonempty and a cost allocation in the core can be obtained from an optimal dual solution [31].

Let I_{ij} be the dual variables associated with constraints (3) and y_i be the dual variables associated with constraints (2), then the dual of the LP relaxation of P is as follows:

$$D: \quad d_L(r) = \max \sum_{(i,j)\in L} r_{ij} I_{ij} \tag{6}$$

$$s.t.\ I_{ij} + y_i - y_j = c_{ij} \quad \forall (i,j) \in L \tag{7}$$

$$y_i - y_j \leq \theta c_{ij} \quad \forall (i,j) \in A \tag{8}$$

$$I_{ij} \geq 0 \quad \forall (i,j) \in L. \tag{9}$$

Let α_{ij} be the allocated cost for covering lane $(i,j) \in L$ and (I^*, y^*) be an optimal solution of D. Then letting $\alpha_{ij} = I^*_{ij}\ \forall (i,j) \in L$ gives a cost allocation in the core and we can compute I_{ij}^* values by solving D in polynomial time. $\square$

Next we show that every core allocation corresponds to an optimal dual solution. First, a necessary and sufficient condition for a given cost allocation to be stable is given in the next lemma.

Lemma 2 *Let α be a cost allocation (not necessarily budget balanced), then α is a stable cost allocation if and only if there exists a vector y that satisfies the linear inequalities*

$$\alpha_{ij} + y_i - y_j \leq c_{ij} \quad \forall (i,j) \in L \quad \textit{and} \quad y_i - y_j \leq \theta c_{ij} \quad \forall (i,j) \in A.$$

Proof: Let $\mathcal{C}^S$ be the set of cycles that cover the lanes in $S \subseteq L$ with minimum total cost z_S^*. Hence, $z_S^* = \sum_{C \in \mathcal{C}^S} z(C)$ where $z(C)$ represents the cost of the cycle C. Also, note that for any cycle C, $\sum_{(i,j) \in C}(y_i - y_j) = 0$. For any α that satisfies the inequalities above, the total cost allocated to any cycle $C \in \mathcal{C}^S$ is less than or equal to the cost of the cycle, since

$$\sum_{(i,j) \in C \cap S} (\alpha_{ij} + y_i - y_j) + \sum_{(i,j) \in C \cap A \setminus S} (y_i - y_j) \leq \sum_{(i,j) \in C \cap S} c_{ij} + \theta \sum_{(i,j) \in C \cap A \setminus S} c_{ij}$$

implying

$$\sum_{(i,j) \in C \cap S} \alpha_{ij} \leq z(C).$$

Hence over all $\mathcal{C}^S$,

$$\sum_{C \in \mathcal{C}^S} \sum_{(i,j) \in C \cap S} \alpha_{ij} \leq z_S^*.$$

The inequality above holds for all $S \subseteq L$, implying that α is a stable cost allocation.

We prove the necessity of the condition by contradiction. Suppose that for a stable cost allocation α, a vector y that satisfies the given linear inequalities does not exist. Then the linear program given below is infeasible.

$$\bar{P} : \max\ 0$$

$$s.t.\ y_i - y_j \leq c_{ij} - \alpha_{ij} \quad \forall (i,j) \in L$$

$$y_i - y_j \leq \theta c_{ij} \quad \forall (i,j) \in A$$

Therefore, the corresponding dual LP below is either infeasible or unbounded.

$$\bar{D} : \min \sum_{(i,j)\in L} c_{ij} x_{ij} + \theta \sum_{(i,j)\in A} c_{ij} z_{ij} - \sum_{(i,j)\in L} \alpha_{ij} x_{ij}$$

$$s.t. \sum_{j\in N} x_{ij} - \sum_{j\in N} x_{ji} + \sum_{j\in N} z_{ij} - \sum_{j\in N} z_{ji} = 0 \quad \forall i \in N$$

$$x_{ij} \geq 0 \quad \forall (i,j) \in L$$

$$z_{ij} \geq 0 \quad \forall (i,j) \in A$$

It is easy to see that $\bar{D}$ is feasible ($x = z = 0$ is a feasible solution). If $\bar{D}$ is unbounded then its optimal objective function value should be equal to $-\infty$. Note that due to the flow balance constraints, any feasible solution to $\bar{D}$ can be decomposed into cycles with positive flows (where this flow is represented by a combination of positive x and z values). In order to achieve an objective function value of $-\infty$, we need to increase the flow over at least one cycle to infinity. Let C be such a cycle and $L_C = C \cap L$ and $A_C = C \cap A$. Then if we increase the flow over cycle C by one, the objective function value is increased by $\sum_{(i,j)\in L_C} c_{ij} + \theta \sum_{(i,j)\in A_C} c_{ij} - \sum_{(i,j)\in L_C} \alpha_{ij}$, which is equal to the cost of the cycle minus the allocated cost of the cycle with a stable cost allocation. By definition of stability, this term is always non-negative for any cycle. Hence the optimal objective function for $\bar{D}$ must be non-negative. Therefore, $\bar{D}$ is neither unbounded nor infeasible implying that $\bar{P}$ is feasible. We conclude that for any stable cost allocation there exists a vector y that satisfies the linear inequalities given above.□

Next, we consider the question whether the set of cost allocations in the core for the shippers' collaboration problem can be completely characterized by the set of optimal solutions to D. That is, whether from each cost allocation in the core, an optimal solution to D can be constructed and vice versa. If this is the case, then additional properties of these cost allocations can be established by studying the set of optimal solutions to D. In related work, Kalai and Zemel [25] prove the coincidence of the core with the set of dual optimal solutions for simple network games, Samet and Zemel [35] extend this result to simple zero-one LP systems, and Engelbrecht-Wiggans and Granot [14] show the same result for a linear production game with a specific structure. None of these settings includes the shipper collaboration game, since we allow multiple players on the arcs. We extend the above results to the flow game described by the LCP with the following corollary that establishes the equivalence of the core and the optimality set of D.

Corollary 1 *The core of the transferable payoffs shippers' collaboration problem is completely characterized by the set of optimal solutions for D.*

Proof: Given Lemma 1, we will only show that every cost allocation in the core corresponds to a feasible solution for D with the optimal objective function value.

Assume that α is a cost allocation in the core. Due to the budget balance property of α and strong duality, $\sum_{(i,j)\in L} r_{ij}\alpha_{ij}$ is equal to the optimal objective function value for D. α satisfies constraints (9) since cost allocations cannot be negative.

Due to Lemma 2, for any stable cost allocation, there exists a vector y that satisfies the linear inequalities $\alpha_{ij} + y_i - y_j \leq c_{ij} \;\; \forall (i,j) \in L$ and $y_i - y_j \leq \theta c_{ij} \;\; \forall (i,j) \in A$. Therefore, for some non-negative matrix ρ, we have;

$$\alpha_{ij} + \rho_{ij} + y_i - y_j = c_{ij} \;\; \forall (i,j) \in L \quad \text{and} \quad y_i - y_j \leq \theta c_{ij} \;\; \forall (i,j) \in A.$$

Therefore, $(\bar{\alpha},\, y)$ where $\bar{\alpha} = \alpha + \rho$ and $\rho \geq 0$ is a feasible solution to D. If any ρ_{ij} is positive, then the optimal objective function value of D would be greater than

$\sum_{(i,j)\in L} r_{ij}\alpha_{ij}$ which either contradicts the budget balance property of α or violates the strong duality property. Therefore, we conclude that $\rho = 0$ and $(\alpha,\ y)$ is an optimal solution for D. Thus, the core is equivalent to the optimality set of D.□

We prove the core of the shippers' collaboration game coincides with the set of optimal dual solutions. Using this result and primal-dual relationships, some properties of the cost allocations in the core can be derived

If, r_{ij}, the number of times that a lane (i,j) must be covered is increased, then cost allocation of that lane, α_{ij}, is expected to increase due to dual sensitivity. Equivalently, the cost allocation of lane (i,j) increases as the marginal benefit of lane (i,j) to the network decreases.

If a lane arc (i,j) is covered more times than it is required (i.e. $z_{ij} > 0$), then its allocated cost is less than the original cost of the lane (i.e. $\alpha_{ij} < c_{ij}$), since by complementary slackness, if $z_{ij} > 0$ then $y_i - y_j = \theta c_{ij}$, and hence $\alpha_{ij} = I_{ij} = (1-\theta)c_{ij}$. Intuitively, if a lane arc is traversed as a deadhead as well, then the contribution of this lane to the collaboration is positive, consequently this lane is charged less than the cost of covering it with a full truckload.

Clearly, any cost allocation constructed from a solution to D satisfies equal treatment of equals property since for each lane (i,j) there is a unique corresponding variable I_{ij}. Moreover, the dummy axiom is satisfied as well. Consider a lane (i,j) with zero marginal contribution to each subset of the collaboration. Adding this lane to the collaborative network (i.e. increasing the respective r_{ij} by one) will not increase the total cost of the collaboration. Then, due to LP duality, the dual optimal objective function value does not change even if a coefficient of a dual variable is increased, which means that the corresponding dual variable, I_{ij}, and the cost allocation must be zero for that lane.

Let the *optimal cycles* be the set of simple cycles obtained by decomposing the optimal solution of the LCP. Note that, the total cost of covering the lanes in L is equal to the sum of the costs of the optimal cycles. Given any optimal cycle, the total cost allocated to the lanes in this cycle is equal to the cost of covering that cycle, since allocating a cost greater than the cost of covering an optimal cycle would make the allocation instable and allocating a smaller cost without increasing the cost of another optimal cycle would not be able to recover the entire budget.

2.4.3 Cost Allocations in the Core

Recall that the nucleolus is the cost allocation that lexicographically maximizes the minimal gain. Intuitively, the goal of finding the nucleolus of a cooperative game with a non-empty core is to find a cost allocation method in the core that avoids favoring any of the players or subsets in the collaboration as much as possible. Clearly, if there is a unique allocation in the core, then that allocation is the nucleolus of the game. Unfortunately, determining the gain for all the subsets for a collaboration explicitly will take exponential time, hence any nucleolus type cost allocation is not efficiently computable. Therefore, we use a similar approach that is appropriate in our context by devising a cost allocation where the allocated costs are proportional to the original lane costs as much as possible.

Let the "*minimum range cost allocation*" be the cost allocation in the core which minimizes the deviation of allocated cost of a lane from its original lane cost over all the lanes in the collaboration when there are alternative optimal solutions in the core of the shippers' collaboration problem. A minimum range core cost allocation can be constructed from a solution to the following LP:

$$
\begin{aligned}
RN: \quad RN(r) = \; & \min k_2 - k_1 & (10)\\
s.t.\ I_{ij} + y_i - y_j = c_{ij} \;\; & \forall (i,j) \in L & (11)\\
y_i - y_j \leq \theta c_{ij} \;\; & \forall (i,j) \in A & (12)\\
I_{ij} \geq 0 \;\; & \forall (i,j) \in L & (13)\\
\sum_{(i,j)\in L} r_{ij} I_{ij} = \; & d_L^*(r) & (14)\\
k_2 c_{ij} \geq I_{ij} \geq k_1 c_{ij} \; & \forall (i,j) \in L. & (15)
\end{aligned}
$$

A solution is feasible to RN if and only if it is an optimal solution for the linear program D. Hence, a cost allocation constructed from an optimal solution to RN finds an allocation in the core that minimizes the percentage deviation between the allocated cost and the original lane cost over all the lanes in the network. Furthermore, the value of k_2 is bounded between 1 and $1+\theta$ and the value of k_1 is bounded between $1-\theta$ and 1. Therefore the optimal objective function value, which gives the "*range*" of the deviation of the cost allocations from the lane costs, is at most 2θ.

This procedure terminates achieving a solution that minimizes the maximum deviation from the original lane costs over all the lanes. Note that this solution may not be a unique optimal solution. Also, different parts of the network may have different characteristics and so this one step algorithm will fail to minimize the maximum deviation value for each part individually. To overcome these problems, an iterative procedure can be devised that uses the LP above as a subroutine and sequentially allocates costs to lanes in different parts of the network. The idea behind this iterative procedure should be to minimize the range for each lane lexicographically to get a unique cost allocation.

2.4.4 Shapley Value

The Shapley Value of a lane (i, j) in the shippers' collaboration game is equal to:

$$\alpha_{ij} = \sum_{S \subseteq L \setminus (i,j)} \frac{| S |! \mid L \setminus (S \cup (i,j)) \mid !}{| L |!} c^{(i,j)}(S) \tag{16}$$

where $c^{(i,j)}(S)$ is the marginal cost of adding lane (i, j) to the subset S. To analyze how the Shapley Value behaves in the shippers' collaboration problem, we consider the simple example from the Introduction section. If only two complimentary lanes are present (i.e. the lanes of shippers A and B), then their Shapley Value is equal to 1. If shipper C enters the network, the total cost of covering the lanes in the network becomes $3 + \theta$ and the Shapley Values become $1 - \frac{\theta}{3}$ for shipper A and $1 + \frac{2\theta}{3}$ for shippers B and C.

Shapley Value is the unique allocation method to satisfy three axioms: dummy, additivity and equal treatment of equals. Although Shapley Value may return cost allocations in the core for some instances of the shippers' collaboration problem, there are many instances where allocations based on Shapley Value are not stable. Furthermore, there exists examples where the cost allocated to a subset of lanes is $(1 + \frac{\theta}{2})$ times the subset's stand alone cost. The allocation in the example is also not stable since $\alpha_A + \alpha_B = 2 + \frac{\theta}{3} > 2$, so Shapley Value is not in the core.

In general, explicitly calculating the Shapley Value requires exponential time. Hence, it is an impractical cost allocation method unless an implicit technique given the particular structure of the game can be found. For the shippers' collaboration problem, the Shapley Value and the core both provide budget balanced but not cross monotonic allocations. However, the core allocations are stable. Given that we also do not know of an efficient technique for calculating the Shapley Value for the shippers' collaboration game, we do not consider it any further in this chapter.

2.4.5 An Important Concept: Cross Monotonicity

In practice, collaborations have a dynamic nature, there may be some new entrants and some members may leave the collaboration over time. Although it is guaranteed that when a new member joins a shippers' collaboration network the overall benefit will be non-negative, some of the existing members' costs may increase if the cost allocation procedures considered do not satisfy the cross-monotonicity property. This is an undesirable situation since some of the existing members then may oppose to an expansion that increases overall benefits. Furthermore, it would be desirable if a contractual agreement can guarantee a new member that the costs initially allocated to its lanes will not increase unless someone leaves the collaboration. Note that since the players for this collaborative game are the lanes and not the shippers, a new member may also correspond to a new lane from an existing shipper.

Next, we discuss whether the allocations in the core are cross monotonic. Unfortunately, we show that it is not possible to find a cost allocation in the core that is also cross monotonic.

Consider the simple example in Figure 2(a), where there are two lanes from node 2 to node 1 belonging to shippers B and C and one lane from node 1 to node 2 belonging to shipper A. Let the cost of covering all the lanes with a truckload be 1 and the deadhead coefficient be θ. If only shippers A and B are present, then the total cost of covering their lanes would be 2. In that case, an allocation in the core must satisfy the following two constraints:

$$\alpha_A + \alpha_B = 2 \quad \text{and} \quad \alpha_A, \alpha_B \leq 1 + \theta.$$

The above constraints imply that in a core cost allocation, the cost allocated to each lane can take on any value in the range $(1 - \theta,\ 1 + \theta)$ as long as the total budget is recovered. If shipper C enters the network, then an allocation in the core must satisfy the following constraints:

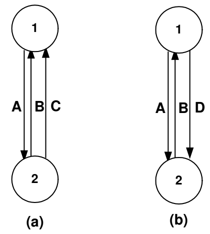

Figure 2: Two shippers' networks.

$$\alpha_A + \alpha_B + \alpha_C = 3 + \theta, \qquad \alpha_A, \alpha_B, \alpha_C \leq 1 + \theta,$$

$$\alpha_A + \alpha_B, \alpha_A + \alpha_C \leq 2, \quad \text{and} \quad \alpha_B + \alpha_C \leq 2 + 2\theta.$$

The dual of the LCP associated with this instance has a unique solution which corresponds to the unique core allocation $\alpha_A = 1 - \theta$ and $\alpha_B = \alpha_C = 1 + \theta$.

Furthermore, if instead of C another shipper, D, with a lane from node 1 to 2 enters the collaboration as in Figure 2(b), then an allocation in the core must satisfy the following constraints:

$$\alpha_A + \alpha_B + \alpha_D = 3 + \theta, \qquad \alpha_A, \alpha_B, \alpha_D \leq 1 + \theta,$$

$$\alpha_A + \alpha_B, \alpha_B + \alpha_D \leq 2, \quad \text{and} \quad \alpha_A + \alpha_D \leq 2 + 2\theta,$$

resulting in the corresponding unique core allocation $\alpha_A = \alpha_D = 1 + \theta$ and $\alpha_B = 1 - \theta$. Now we can conclude that no matter which one of the core allocations for shippers A and B were chosen initially, an adversary can present either shipper C or D to join the network increasing one of the initial allocations. Hence even for this simple instance of the shippers' collaboration problem there does not exist a cross monotonic cost allocation in the core.

Although this result is discouraging, it is also expected due to the nature of the shippers' collaboration problem. A new member joining the collaboration can create many alternative tours (cycles) for the cost minimization problem. While the network as a whole may enjoy these alternatives, there may be some individual shippers disadvantaged with the new situation. Since, a newcomer's effect on the collaboration entirely depends on the initial network structure, having a cost allocation method that benefits all the existing shippers and is also budget balanced is impossible.

As expected, a cost allocation method cannot satisfy all the desired properties for the shippers' collaboration problem. Hence, alternative cost allocation methods will be of value when different aspects of a successful collaboration are important. In the following, we discuss relaxing the budget balanced condition to design a cross monotonic cost allocation method.

2.4.5.1 A Budget Recovery Bound for Cross Monotonic and Stable Allocations

Since there does not exist a budget balanced and cross monotonic cost allocation for the shippers' collaboration problem, we next study cross monotonic and stable allocations that recover a good percentage of the total cost. Note that, allocating costs greater than the total budget of the collaboration is not relevant in this context since in such an allocation the stability condition will be violated as well. First, we show that there exists a worst case bound on the percentage of the total budget that any cross monotonic and stable cost allocation method can recover. Then, we present a simple cost allocation method that recovers at least the same percentage for every instance of the problem.

Lemma 3 *A cross monotonic and stable cost allocation in which the total allocated cost is always less than or equal to the total cost of the collaboration can guarantee to recover at most* $\frac{1}{1+\theta}$ *of the total cost.*

Proof: Consider the example in Figure 2 where only shippers A and B are present. Let α_{12} and α_{21} be the cost allocated to shippers A and B respectively with a cross monotonic and stable cost allocation. Note that the total cost allocated to both shippers cannot be greater than the total cost of the collaboration, hence $\alpha_{12}+\alpha_{21} \leq 2$

due to the stability restriction. If the number of times lane $(2,1)$ must be covered, that is r_{21}, is increased, then the total cost of the collaboration goes to $(1+\theta)r_{21}$ as r_{21} approaches infinity. The fraction of the total budget recovered with a cross monotonic and stable cost allocation is less than or equal to $\frac{\alpha_{12}+r_{21}\alpha_{21}}{(1+\theta)r_{21}}$. Similarly, if r_{12} is increased, then the total cost of the collaboration goes to $(1+\theta)r_{12}$ as r_{12} approaches infinity and the budget recovery fraction is less than or equal to $\frac{r_{12}\alpha_{12}+\alpha_{21}}{(1+\theta)r_{12}}$. In order to guarantee the best fraction of the total budget recovered, we need to maximize the minimum fraction $(\max\{\min\{\frac{\alpha_{12}+r_{21}\alpha_{21}}{(1+\theta)r_{21}}, \frac{r_{12}\alpha_{12}+\alpha_{21}}{(1+\theta)r_{12}}\}\})$, therefore both lanes should have a cost allocation equal to 1, $(\alpha_{12} = \alpha_{21} = 1)$. Hence, any cross monotonic and stable cost allocation method cannot recover more than $\frac{1}{1+\theta}$ fraction of the total budget when either r_{21} or r_{12} approaches infinity.

This example represents a worst case instance since there exists a cross monotonic and stable cost allocation that guarantees to recover at least $\frac{1}{1+\theta}$ fraction of the total budget for every instance of the problem. The simple stable cost allocation method where every shipper pays its own lane cost and no one pays for the deadhead costs of the network (i.e. $\alpha_{ij} = c_{ij}$) is cross monotonic, since allocated cost of a lane is independent of other lanes. The amount of recovered budget is equal to $\sum_{(i,j)\in L} c_{ij}$ with this simple allocation mechanism. The total cost of covering all the lanes in the network is composed of two parts: total cost of the lanes and the total deadhead cost. It is easy to see that the total deadhead cost is less than or equal to θ times the total cost of the lanes, $\theta \sum_{(i,j)\in L} c_{ij}$. Then, this simple stable cost allocation method recovers at least $\frac{1}{1+\theta}$ fraction of the total budget for every instance. Therefore, we conclude that a cross monotonic and stable cost allocation can guarantee to recover at most $\frac{1}{1+\theta}$ of the total budget of the collaboration. □

Immorlica et al. [23] study various problems such as edge cover, vertex cover, set cover, and metric facility location and presents the limitations of cross monotonic cost allocations in budget recovery. Lemma 3 proves that the same limitations hold for the shippers' collaboration game as well.

2.4.5.2 A Cross Monotonic and Stable Cost Allocation Method

In the above subsection, we developed an upper bound on the guaranteed fraction of the budget recovered by any cross monotonic and stable cost allocation. We also pointed out a simple cost allocation method which guarantees this upper bound for all instances. However, this simple allocation is ineffective for recovering the best possible fraction of the budget in most instances. Therefore, we attempt to find a cross monotonic and stable cost allocation method that guarantees the same bound and may perform better in some instances.

Recall the minimum range core allocation developed in Subsection 2.4.3 which minimizes the deviation between the allocated cost and the original lane cost over all the lanes in the network. By using a similar idea, we develop a procedure that finds cross monotonic and stable cost allocations while maximizing the amount of budget recovered. Cross monotonicity of the allocation is ensured by allocating a cost that is a certain percentage of the original lane cost to each lane when the lane joins the collaboration and not allowing to increase this percentage afterwards. Let k be the variable denoting this percentage. (Note that k may take values greater than 1.) In this procedure, CM, a modified version of the linear program D is solved every time a new lane enters the collaboration and an optimal solution (I^*, y^*, k^*) is obtained. Then α_{ij}, the cost allocated to lane $(i,j) \in L$ is set equal to $I^*_{i,j}$ for all (i,j), recovering $d_L^{CM*}(r)$ of the total cost.

$$CM : d_L^{CM}(r) = \max \sum_{(i,j)\in L} r_{ij} I_{ij} \tag{17}$$

$$s.t.\ I_{ij} + y_i - y_j = c_{ij} \quad \forall (i,j) \in L \tag{18}$$

$$y_i - y_j \leq \theta c_{ij} \quad \forall (i,j) \in A \tag{19}$$

$$I_{ij} = k c_{ij} \quad \forall (i,j) \in L. \tag{20}$$

Corollary 2 *The procedure described above provides a cross monotonic and stable cost allocation method which guarantees to recover at least $\frac{1}{1+\theta}$ of the total budget.*

Proof: Due to constraints (18)-(19), a cost allocation $\alpha_{ij} = I^*_{ij}$ where I^*_{ij} is an optimal solution to CM is stable. Also, due to constraints (20), the objective function is equal to $d_L^{CM}(r) = \max \sum_{(i,j)\in L} r_{ij} I_{ij} = \max k \sum_{(i,j)\in L} r_{ij} c_{ij}$. Hence CM may be solved with the objective $\max k$ because $\sum_{(i,j)\in L} r_{ij} c_{ij}$ is a constant.
It is easy to see that $I_{ij} = c_{ij}\ \forall (i,j) \in L$, $y_i = 0\ \forall i \in N$, and $k = 1$ is a feasible solution to CM. Therefore, a feasible solution to CM provides a cost allocation that covers at least $\sum_{(i,j)\in L} r_{ij} c_{ij}$. Recall that this value guarantees $\frac{1}{1+\theta}$ fraction of the total budget is recovered.
Next we show that the above procedure provides a cross monotonic cost allocation. Consider adding a new shipper with lane (i,j) to the collaboration. If (i,j) is already in L, the constraint set of CM remains the same and the optimal k value and the costs allocated to lanes do not change. If (i,j) is not in L, then additional constraints (stability conditions and the constraint $I_{ij} = kc_{ij}$) are introduced which may further restrict the k value. Since CM's feasible region becomes smaller, the value of k (and so the allocated cost of the lanes) does not increase with additional members in the collaboration.□

This method in general is expected to recover a greater fraction of the total budget than allocating only the lane costs, however the budget recovery is still bounded with $\frac{1}{1+\theta}$ in the worst case. We believe that allocation methods with such recovery bounds are impractical and will not be used in the industry. However, due to the nature of the problem and the cross monotonicity property, no cost allocation method can do any better.

Another drawback of this method is observed when a cycle consists of only the lane arcs. Recall that the total cost allocated to a cycle should be less than or equal to the cost of the cycle due to the stability condition. In that case, the procedure above yields $k^* = 1$, which means that every lane in the collaboration is allocated only the lane costs and no one pays for the deadhead costs. In a large collaboration with a considerable number of lanes, this phenomena is most likely to happen as we observed in our computational study. Similar to the minimum range core allocation developed in Subsection 2.4.3, an iterative procedure that sequentially allocates costs to the lanes in different parts of the network may be employed to eliminate this drawback.

2.5 Cost Allocation Methods with Additional Properties

In the previous sections, we have considered the standard cost allocation properties from cooperative game theory and proved that there does not exist a cost allocation method that is in the core and cross monotonic. However, for the shippers' collaboration problem there are other desirable cost allocation properties besides these well-known concepts. A shipper joins a collaborative network to reduce its total transportation expenses by reducing asset repositioning costs incurred while covering its lanes with a truckload. Hence a collaborative transportation procurement network should not allocate the entire stand alone cost or a cost that is less than the truckload lane cost to any shipper. We will first consider a "minimum liability" cost allocation where every lane pays at least its original truckload lane cost and then a "positive benefit" cost allocation where every lane is ensured to have a certain discount from its stand alone cost.

In following subsections, we attempt to develop methods that provide cost allocations satisfying these new properties while violating the formerly discussed properties minimally. That is, we first relax either budget balance or stability conditions and impose the restrictions associated with the new properties, and then attempt to find an allocation with a minimum deviation from the relaxed property.

2.5.1 Cost Allocation Methods with the Minimum Liability Restriction

In the previous sections, we considered cost allocations with transferable payoffs where the allocated cost of some of the lanes may be less than their original lane cost. In this context, not only the asset repositioning costs but also the total lane covering costs are distributed among all the shippers without any restrictions leading to the existence of "free-riders" with zero allocated costs.

In a real world situation, shippers may not be pleased to cover some other shipper's truckload expense and to have a free-rider in the collaboration. Therefore, we next develop a cost allocation method where shippers are responsible at least for their truckload lane costs and only the total deadhead cost is distributed among the members of the collaboration. First, we show that there may be no solution in the core that allocates at least the original lane cost to each lane.

Lemma 4 *The core of the shippers' collaboration problem may not include a cost allocation that satisfies the minimum liability restriction.*

Proof: We provide an instance of the problem with no such cost allocation. Consider the simple example from the Introduction section where shipper A has a lane from node 1 to node 2 and shippers B and C each have lanes from node 2 to node 1. The cost of covering each lane with truckload is equal to 1 and as a deadhead is θ. In a minimum liability cost allocation, $\alpha_{12} \geq 1$ and $\alpha_{21} \geq 1$. If $\alpha_{12} = \alpha_{21} = 1$, then the cost allocation is not budget balanced. If either α_{12} or α_{21} is greater than 1, then the cost allocation is not stable. Hence the cost allocations in the core for the above instance of the shippers' collaboration problem do not satisfy the minimum liability restriction.$\square$

To determine whether the core of a given instance of the shippers' collaboration problem includes a cost allocation with the minimum liability restriction, the linear program D may be solved with the additional constraint $I_{ij} \geq c_{ij} \ \ \forall (i,j) \in L$. If the objective function value for this constrained version of D is equal to the total cost of covering all the lanes, then we conclude that there exists a solution in the core satisfying the minimum liability restriction.

As the cost allocations in the core may not satisfy the minimum liability restriction, we may consider relaxing the budget balance or the stability conditions to design cost allocation methods with the minimum liability restriction.

2.5.1.1 Cost Allocations with a Maximum Budget Recovery Percentage

Finding a cost allocation method with the minimum liability restriction may require us to relax either the budget balance or the stability properties. Since it is easier to deal with the budget balance property than the stability property, we consider it first. Although a cost allocation method ideally recovers the total cost of the collaboration, it may be still worthwhile to investigate budget balanced relaxed cost allocation methods. However, any budget balanced relaxed cost allocation method has a lower bound on the fraction of the budget that can be recovered that is far from promising for its use in practice. A minimum liability cost allocation with the best possible budget recovery can be found by solving the following LP:

$$BB : d_L^{BB}(r) = \max \sum_{(i,j)\in L} r_{ij} I_{ij} \tag{21}$$

$$s.t.\ I_{ij} + y_i - y_j \le c_{ij} \quad \forall (i,j) \in L \tag{22}$$

$$y_i - y_j \le \theta c_{ij} \quad \forall (i,j) \in A \tag{23}$$

$$I_{ij} \ge c_{ij} \quad \forall (i,j) \in L. \tag{24}$$

Lemma 5 *The cost allocation obtained by letting $\alpha_{ij} = I^*_{ij}\ \forall (i,j) \in L$, where I^* is the optimal solution for the linear program BB, provides a stable cost allocation for the shippers' collaboration problem with the minimum liability restriction that recovers* *the maximum possible percentage of the total budget. The lower bound on the fraction of total cost recovered is* $\frac{1}{1+\theta}$.

Proof: Due to Lemma 2 and constraints (24), any stable minimum liability cost allocation is a feasible solution to BB. Hence the optimal solution to BB provides an allocation that maximizes the total budget recovered among all the stable minimum liability cost allocations. Since $\sum_{(i,j)\in L} r_{ij}\alpha_{ij} = \sum_{(i,j)\in L} r_{ij}I^*_{ij} \geq \sum_{(i,j)\in L} r_{ij}c_{ij}$, we conclude that the second part of the lemma is satisfied as well. Note that this lower bound is tight for the example discussed in the proof of Lemma 3.□

2.5.1.2 Cost Allocations with a Minimum Percentage Deviation from Stability

Stability is the key property that holds a collaboration together and without this property a collaboration is under the risk of collapsing. On the other hand, in many real life situations, this threat is not as immediate as we conclude from a theoretical perspective. Even if there exists a sub-coalition that is better off by collaborating on its own, this usually does not mean that the grand coalition will break down. In real life, due to insufficient information sharing, limited rationality of the players, cost of collaborating, and contractual agreements many of the sub-coalitions are not formed even though they create additional benefits for its members. Therefore, relaxing the stability conditions in a limited way can be justified.

Relaxing the stability condition in order to find a budget balanced minimum liability cost allocation is not as straightforward as relaxing the budget balance property. There is an exponential number of subsets to consider while seeking an allocation method which has minimum deviation from stability over all these subsets. For any subset of the collaboration, the deviation from stability can be defined as the percent deviation or the fixed value deviation from stand alone cost of the subset. We next develop a procedure that finds cost allocations with minimum deviation from stability.

In the method we propose, the objective is to minimize the maximum percentage deviation of the allocated cost of a subset from its stand alone cost over all the subsets of the collaboration while making sure that the total budget is recovered and the allocated cost of a lane is greater than or equal to its original lane cost. This method is a quick method (computation times are provided in Section 2.7) but most importantly it provides a minimum liability cost allocation with minimum possible percentage deviation.

First, we show that by generalizing Lemma 2, we can calculate the maximum percentage deviation from stability of a given cost allocation efficiently, that is without considering an exponential number of lane subsets.

Theorem 1 *Let α be a cost allocation. For a given scalar, k, if there exists a vector y that satisfies the linear inequalities*

$$\alpha_{ij} + y_i - y_j \leq c_{ij}(1+k) \quad \forall (i,j) \in L \quad \textit{and} \quad y_i - y_j \leq \theta c_{ij}(1+k) \quad \forall (i,j) \in A,$$

then the deviation from stability for α over all subsets is at most k. Let k^* be the minimum of such k values, then there exists a subset $S \subseteq L$ such that $(1+k^*)z_S^* = \sum_{(i,j)\in S} \alpha_{ij}$.

Proof: As in the proof of Lemma 2, one can easily prove that

$$\sum_{C \in \mathcal{C}^S} \sum_{(i,j)\in C\cap S} \alpha_{ij} \leq z_S^*(1+k),$$

implying that the deviation from stability for α is at most k for S. Since S is an arbitrary subset of L, we conclude that the deviation from stability for α over all subsets is at most k, hence the percentage deviation is $100 \times k$.

Next, we prove by contradiction that there exists a subset $S \subseteq L$ such that $(1+k^*)z_S^* = \sum_{(i,j)\in S} \alpha_{ij}$. More specifically, we show that there exists a cycle C such that $(1+k^*)z(C) = \sum_{(i,j)\in C\cap L} \alpha_{ij}$. Consider the primal dual LP pair given below for a cost allocation α and scalar k.

$$\tilde{P}(\alpha, k) : \max\ 0$$

$$s.t.\ y_i - y_j \leq c_{ij}(1+k) - \alpha_{ij} \quad \forall (i,j) \in L$$

$$y_i - y_j \leq \theta c_{ij}(1+k) \quad \forall (i,j) \in A.$$

$$\tilde{D}(\alpha, k) : \min (1+k) \sum_{(i,j)\in L} c_{ij}x_{ij} + (1+k)\theta \sum_{(i,j)\in A} c_{ij}z_{ij} - \sum_{(i,j)\in L} \alpha_{ij}x_{ij}$$

$$s.t. \sum_{j\in N} x_{ij} - \sum_{j\in N} x_{ji} + \sum_{j\in N} z_{ij} - \sum_{j\in N} z_{ji} = 0 \quad \forall i \in N$$

$$x_{ij} \geq 0 \quad \forall (i,j) \in L$$

$$z_{ij} \geq 0 \quad \forall (i,j) \in A.$$

Note that for a given allocation α and its corresponding k^*, both $\tilde{P}(\alpha, k^*)$ and $\tilde{D}(\alpha, k^*)$ are feasible and bounded due to definition of k^*. Now suppose that for α and k^* all the cycles satisfy

$$(1+k^*) \sum_{(i,j)\in C\cap L} c_{ij} + (1+k^*)\theta \sum_{(i,j)\in C\cap A} c_{ij} > \sum_{(i,j)\in C\cap L} \alpha_{ij}$$

that is

$$(1+k^*)z(C) > \sum_{(i,j)\in C\cap L} \alpha_{ij}.$$

Therefore, for some $\bar{k} < k^*$, any cycle satisfies

$$(1+\bar{k})z(C) \geq \sum_{(i,j)\in C\cap L} \alpha_{ij}.$$

Note that $\tilde{D}(\alpha, \bar{k})$ is feasible. Furthermore, due to the above inequality and similar arguments as in the proof of Lemma 2, by increasing flow over any cycle, we cannot decrease the optimal objective function of $\tilde{D}(\alpha, \bar{k})$ below 0. Hence, $\tilde{D}(\alpha, \bar{k})$ is also bounded. Consequently, $\tilde{P}(\alpha, \bar{k})$ is also feasible and bounded contradicting with the minimality of k^*. Therefore we conclude that there exists at least one cycle C such that the cost allocated to cycle C is equal to $(1 + k^*)$ times the stand-alone cost of cycle C, $(1 + k^*)z(C) = \sum_{(i,j)\in C\cap L} \alpha_{ij}$. $\square$

The second part of Theorem 1 shows that k^* is not just an upper bound on the deviation from stability over all subsets, but it also represents the exact deviation from stability of a given cost allocation. For a given cost allocation, the maximum deviation from stability, k^*, can be determined by solving a linear program with the objective of minimizing k over the linear inequalities given above without considering an exponential number of lane subsets.

Next, we describe the different steps of the cost allocation method we propose to minimize the deviation with regard to stability while ensuring minimum liability and budget-balance properties.

<u>Method SR:</u>

Step 1: Find the minimum total cost of covering all the lanes in the network by solving either P or D. Let this cost be $z_L^*(r)$.

Step 2: Using the solution in Step 1, formulate and solve the LP below:

$$SR: \ d_L^{SR}(r) = \min k \tag{25}$$

$$s.t. \sum_{(i,j)\in L} r_{ij} I_{ij} = z_L^*(r) \tag{26}$$

$$I_{ij} \geq c_{ij} \quad \forall (i,j) \in L \tag{27}$$

$$I_{ij} + y_i - y_j \leq (1+k)c_{ij} \quad \forall (i,j) \in L \tag{28}$$

$$y_i - y_j \leq \theta c_{ij}(1+k) \quad \forall (i,j) \in A. \tag{29}$$

Step 3: Let (I^*, y^*, k^*) be an optimal solution to SR. Then let $\alpha_{ij} = I_{ij}^*$ $\forall (i,j) \in L$.

Theorem 2 *The procedure above gives a budget balanced cost allocation with the minimum liability restriction where the optimal k value, k^*, is the maximum deviation from stability over all the subsets of the collaboration. k^* represents the minimum possible deviation from stability over all the budget balanced cost allocations with the minimum liability restriction. The value of k^* is bounded from above by $\frac{\theta}{2}$ which, in fact, is the best upper bound.*

Proof: By construction SR provides a budget balanced cost allocation that satisfies the minimum liability restriction. Also due to the second part of Theorem 1, the maximum deviation of allocated cost from the stand alone cost over all subsets is exactly equal to k^*.

For the second part of the theorem, we note that all the cost allocations that satisfy the minimum liability and budget balanced restrictions are feasible for SR. Hence due to Theorem 1, the cost allocation found by Method SR will have the minimum possible deviation from stability over all the budget balanced cost allocations with the minimum liability restriction.

For the last part of the theorem, we first construct a feasible solution to the linear program SR from the optimal solution to D with a k value equal to $\frac{\theta}{2}$. By doing so, we prove that SR is always feasible and k^* is bounded from above by $\frac{\theta}{2}$, since SR is a minimization problem.

Let $(\bar{I}, \bar{y})$ be the optimal solution to D, the dual of the lane covering problem. Due to the constraints (7) of model D, $\bar{I}_{ij} \geq c_{ij}$ if $(\bar{y}_j - \bar{y}_i)$ is nonnegative, otherwise $\bar{I}_{ij} < c_{ij}$. Let's partition the lane set into two sets: $L_1 = \{(i,j) : \bar{I}_{ij} \geq c_{ij}\}$ and $L_2 = \{(i,j) : \bar{I}_{ij} < c_{ij}\}$.

Now, we construct a feasible solution to SR from $(\bar{I}, \bar{y})$. Let $\hat{I}_{ij}$ and $\hat{y}_i$ be such that:

$$\hat{I}_{ij} = \bar{I}_{ij} - \rho_{ij} \geq c_{ij} \quad \forall (i,j) \in L_1,$$

$$\hat{I}_{ij} = c_{ij} \quad \forall (i,j) \in L_2,$$

$$\hat{y}_i = \frac{\bar{y}_i}{2} \quad \forall i$$

where ρ_{ij}'s are selected so that $\sum_{(i,j)\in L} r_{ij}\hat{I}_{ij} = z_L^*(r)$. Note that it is possible to find such $\rho \geq 0$ because $\sum_{(i,j)\in L} r_{ij}c_{ij} \leq z_L^*(r)$. Also, note that $\hat{y}_i - \hat{y}_j = \frac{\bar{y}_i}{2} - \frac{\bar{y}_j}{2} \leq \frac{\theta}{2}c_{ij}$ $\forall (i,j) \in A$ due to constraint (8) of D and $c_{ij} = c_{ji}$ $\forall (i,j) \in L$ (Euclidian graph).

We claim that $(\hat{I},\, \hat{y},\, \frac{\theta}{2})$ is a feasible solution to SR. Constraints (26) and (27) are satisfied due to steps described above to obtain $(\hat{I},\, \hat{y})$. Constraints (29) are satisfied since

$$\hat{y}_i - \hat{y}_j \leq \frac{\theta}{2}c_{ij} < \theta c_{ij}(1 + \frac{\theta}{2}) \quad \forall (i,j) \in A.$$

Constraints (28) are partitioned into two sets with respect to L_1 and L_2 and are satisfied as shown below since $\hat{y}_i - \hat{y}_j \leq \frac{\theta}{2}c_{ij}$ $\forall (i,j) \in A$, $c_{ij} = c_{ji}$ $\forall (i,j) \in L$ and $(\bar{I}, \bar{y})$ is the optimal solution to D.

$$\hat{I}_{ij} + \hat{y}_i - \hat{y}_j = \bar{I}_{ij} - \rho_{ij} + \hat{y}_i - \hat{y}_j = c_{ij} + \bar{y}_j - \bar{y}_i - \rho_{ij} + \hat{y}_i - \hat{y}_j$$

$$= c_{ij} + \hat{y}_j - \hat{y}_i - \rho_{ij} \leq c_{ij}(1 + \frac{\theta}{2}) \quad \forall (i,j) \in L_1 \quad \text{and}$$

$$\hat{I}_{ij} + \hat{y}_i - \hat{y}_j = c_{ij} + \hat{y}_i - \hat{y}_j \leq c_{ij}(1 + \frac{\theta}{2}) \quad \forall (i,j) \in L_2.$$

Thus, we conclude that $(\hat{I},\, \hat{y},\, \frac{\theta}{2})$ is a feasible solution and the optimal objective function value of SR is bounded above by $\frac{\theta}{2}$.

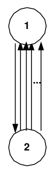

Figure 3: Infinity Example.

Moreover, $\frac{\theta}{2}$ is the best upper bound, that is, there exists an instance where the optimal value of k is equal to $\frac{\theta}{2}$. Consider the example in Figure 3 where there is one lane from node 1 to node 2 and an infinite number of lanes from node 2 to node 1. For this instance, the unique budget balanced minimum liability cost allocation allocates a cost of 1 to the lane from node 1 to 2 and $1+\theta$ to each lane from node 2 to 1. Given these values, cost allocated to each cycle consisting of arcs $(1, 2)$ and $(2, 1)$ is equal to $2 + \theta$ which deviates by $\frac{\theta}{2}$ from the stand alone cost of the corresponding subset of lanes. $\square$

2.5.2 Cost Allocation Methods with Guaranteed Positive Benefits

A collaboration is established to create benefits for its members, therefore any given shipper expects to have positive gains when entering a shippers' collaboration. We next discuss cost allocation methods, which ensure that each shipper is charged less than its stand alone cost or equivalently that each lane in the collaborative network has a "positive benefit". Similar to the cross monotonicity property, such methods are desirable in contracting stages since promising each player at least a certain reduction from its stand alone costs increases the attractiveness of the collaboration.

As we have presented previously there exists instances of the shippers' collaboration problem where all the cost allocations in the core allocates the stand alone costs to some of the shippers. In such situations, there exists a set of shippers who have no gains from being in the collaboration. This might lead the shippers to leave the collaboration (or at least move out some of their lanes from the collaborative network) to avoid costs and problems associated with coordinating with other members. It may not be possible to find a budget balanced and stable cost allocation if we impose that the allocation should also provide a positive benefit to each member. Hence, we study positive benefit cost allocation mechanisms where one of these conditions is relaxed.

For instances where there are very limited or no synergies among the lanes of a collaborative network, a cost allocation with positive benefits to its members will not be possible without a budget deficit as should be expected. In such situations, the collaborations are not expected to survive.

Offering any shipper an allocated cost less than its stand alone cost may not be enough for the stability of the collaboration, since there may be other possibilities for the shipper to exploit, such as sub-coalitions, offering even higher benefits. However, it is not unreasonable to assume that the members of a collaborative network have limited rationality or equivalently each shipper cannot be aware of the structure of the entire network in order to exploit it. Hence, a collaboration ensuring a certain percentage reduction from the stand alone costs may still be stable.

2.5.2.1 Relaxing the Budget Balance Property

First, we develop a stable cost allocation method with positive benefits while relaxing the budget balance property. Let $0 < \sigma < 1$ denote the maximum fraction of the stand alone cost allowed to be allocated to each lane. That is, in a resulting cost allocation, each lane will have at least $100 \times (1 - \sigma)$ percent savings from its stand alone cost. Note that, we assume σ is pre-specified by the collaboration. Let (I^*, y^*) be an optimal solution to the following linear program B. Then a stable allocation α can be constructed by letting $\alpha_{ij} = I^*_{ij}\ \forall (i,j) \in L$, where the allocated cost of any lane is at most $100 \times \sigma$ percent of its stand alone cost and the fraction of the budget recovered is maximized.

$$B : d_L^{BBSA}(r, \sigma) = \max \sum_{(i,j) \in L} r_{ij} I_{ij} \tag{30}$$

$$s.t.\ I_{ij} + y_i - y_j \leq c_{ij} \quad \forall (i,j) \in L \tag{31}$$

$$y_i - y_j \leq \theta c_{ij} \quad \forall (i,j) \in A \tag{32}$$

$$I_{ij} \leq (1 + \theta) c_{ij} \sigma \quad \forall (i,j) \in L. \tag{33}$$

Next we try to find an upper bound for the budget deficiency of the cost allocation constructed with this method. Let $\hat{\alpha}$ be a cost allocation in the core for the shippers' collaboration problem. Let L_c be the set of lanes such that $\hat{\alpha}_{ij} > (1+\theta)c_{ij}\sigma$. If we update the allocation of the lanes in L_c by decreasing their allocations so that we get a feasible solution for B, then the budget deficiency of this allocation will be at most

$$\sum_{(i,j)\in L_c} r_{ij}c_{ij}(1+\theta)(1-\sigma) \leq \sum_{(i,j)\in L} r_{ij}c_{ij}(1+\theta)(1-\sigma) = (1-\sigma)\sum_{(i,j)\in L} a_{ij}$$

where $a_{ij} = (1+\theta)c_{ij}$ is the stand alone cost of lane (i,j). If the pre-specified σ values are different for different lanes then the budget deficiency of this allocation will be at most

$$\sum_{(i,j)\in L} r_{ij}c_{ij}(1+\theta)(1-\sigma_{ij}) = \sum_{(i,j)\in L} (1-\sigma_{ij})a_{ij}.$$

Note that this bound is tight for a collaborative network with no synergies between its lanes. For the above constructed allocations the remaining fraction of the budget not recovered may be too large to be distributed to the members of the collaboration in other ways. Therefore, we consider relaxing the stability property next.

2.5.2.2 Relaxing the Stability Property

Now we develop a method, which provides a positive benefits allocation that recovers the entire budget and has minimum percentage instability over all the subsets of the collaboration. We use a similar methodology to Method SR of Subsection 2.5.1.2 and first solve the following linear program:

$$SS: d_L^{SSA}(r,\sigma) = \min k \tag{34}$$

$$s.t.\ I_{ij} + y_i - y_j \leq c_{ij}(1+k) \quad \forall (i,j) \in L \tag{35}$$

$$y_i - y_j \leq \theta c_{ij}(1+k) \quad \forall (i,j) \in A \tag{36}$$

$$\sum_{(i,j)\in L} r_{ij}I_{ij} = z_L^*(r) \tag{37}$$

$$I_{ij} \leq (1+\theta)c_{ij}\sigma \quad \forall (i,j) \in L. \tag{38}$$

As before we assume that σ is pre-specified. If we let $\alpha_{ij} = I^*_{ij} \ {}_{\forall(i,j)\in L}$ where (I^*, y^*, k^*) is an optimal solution to SS, then α is a budget balanced allocation which allocates at most $100 \times \sigma$ percent of its stand alone cost to each lane and the maximum instability over all the subsets of the collaboration is $100 \times k^*$ percent. However in this case SS might be infeasible. In fact, for collaborative networks with little synergy between its lanes the allocation problem might be infeasible even for a small σ value. A rough lower bound for σ which ensures the feasibility of SS is equal to $\frac{z^*_L(r)}{\sum_{(i,j)\in L} a_{ij}}$. The optimal k depends on the pre-specified σ value and the instance considered.

2.6 Cost Allocations Under Different Carrier Cost Structures

Throughout the chapter we have assumed that the carrier's cost structure is such that if c_{ij} is the cost of traversing an arc with a full truckload then the cost of traversing the same arc as a deadhead is equal to θc_{ij} for some $0 < \theta \leq 1$. In this section, we present two different cost structures that might be used by the carriers in the industry and discuss how these affect the allocation methods described above.

2.6.1 Multiple Deadhead Coefficients

The deadhead costs over the arcs of a carrier's network may not be homogenous and the associated cost of traversing an arc (i, j) might be equal to $\theta_{ij} c_{ij}$. Such a cost structure captures the fact that different parts of the network have different characteristics such as load imbalance.

The existence of multiple deadhead coefficients does not affect much of the procedures described above. First of all, the LCP can be solved with the same minimum cost circulation linear program by replacing θ with θ_{ij} appropriately. When D, the dual of the LP relaxation of the LCP is considered, the only change occurs in the constraints $y_i - y_j \leq \theta c_{ij} \ \forall (i,j) \in A$. Updating this constraint set with $y_i - y_j \leq \theta_{ij} c_{ij} \ \forall (i,j) \in A$ and solving the dual problem provides a cost allocation in the core. Thus, all the procedures described above remain the same, however, the bounds should be updated accordingly.

2.6.2 Fixed Deadhead Costs

A carrier might charge a fixed cost for each deadhead rather than a variable cost based on the truckload cost of the arc. Let θ_f be this fixed deadhead charge. In this case, the LCP is modified by replacing the objective function with

$$\min \sum_{(i,j)\in L} c_{ij}x_{ij} + \theta_f \sum_{(i,j)\in A} z_{ij}.$$

Accordingly the dual of the LCP is modified by replacing the constraints (8) with $y_i - y_j \leq \theta_f \ \forall (i,j) \in A$. Note that the "fixed deadhead cost" problem can be converted to the "multiple deadhead cost" problem by letting $\theta_{ij} = \frac{\theta_f}{c_{ij}}$. Then the constraints $y_i - y_j \leq \theta_f \ \forall (i,j) \in A$ are replaced with $y_i - y_j \leq \theta_{ij}c_{ij} \ \forall (i,j) \in A$ and the rest follows. Thus, the procedures remain the same and the bounds should be modified.

2.7 Computational Study

We have carried out a computational study to evaluate the performance of the methods we have discussed on randomly generated instances and real-life instances. The real-world data is provided by a strategic sourcing consortium for a $14 billion dollar sized US industry. The company name is kept confidential at the company's request. Although we presented theoretical bounds for the key performance metrics for most of the allocation methods discussed above, we believe performance of these methods will be significantly better on average.

The random instances we use are generated within a 1,000 $\times$ 1,000 square with different number of nodes, average number of edges per node, number of clusters and ratio of cluster points to total points. The first two parameters control the size of the instance generated. Clusters represent the dense regions with many supply or demand points such as metropolitan areas and the cluster ratio controls how many of the total points fall within a cluster. Clusters are uniformly distributed in the square and each point in a cluster is generated by using a standard Normal distribution. Remaining points are distributed uniformly across the map. We do not generate any lanes between any two points of a cluster.

Finally, we generate a set of instances simulating a supply chain structure. In a "supply chain instance", all the points generated are divided into three categories : Suppliers, plants and distribution centers, and customers (with 0.2, 0.1, 0.7 fraction of the points) and lanes are directed from suppliers to plants and DC's and from plants and DC's to customers. Lanes among plants and DC's are also allowed. The supply chain instances are more relevant when the collaboration is intra-firm where as the random instances represent an inter-firm collaborative network.

We generate instances with 50, 100, 200 and 500 points, 2, 5 and 10 average lanes per point, 5 and 10 clusters and 0, 0.3 and 0.6 cluster ratios. For each combination of these parameters, we generate two random instances, with and without a supply chain structure. We perform each experiment with the deadhead coefficients 0.25, 0.5, 0.75 and 1. In order to investigate the effect of each parameter on the key performance indicators, we take the average of the results on the 576 instances generated with respect to these parameters. These results are presented in Table 1 and Table 2. We remind the reader that there exist a drawback of the cross monotonic and stable cost allocation method (CM). Recall that when a cycle consists of only lane arcs, the procedure CM yields a cost allocation where every lane in the collaboration is allocated only the lane costs and this phenomena is most likely to happen in large instances. Thus, the fraction of the budget recovered with the CM method is not a meaningful measure of performance for large instances and because of this we do not present computational results on the CM method.

First, most of our procedures are very efficient and complete within a few minutes. For example, for the largest instances tested with 500 nodes the average solution times for the procedures primal (P), dual (D), minimum range cost allocation (RN), cross monotonic cost allocation (CM), budget recovery (BB), and Method SR (SR) are 20.42, 50.68, 63.35, 9.37, 37.73 and 149.15 seconds respectively. The solution times of the procedures with the positive benefits restriction highly depend on the pre-specified σ and θ values and the characteristics of the particular instance.

In Table 1, the column "DH %" presents the percentage of the total deadhead cost to the total cost of the instance. The column"Range" represents the optimal objective function value of the core cost allocation method discussed in Subsection 2.4.3, which measures the range of the deviation of the cost allocations from the lane costs. In order to make the average value of this performance metric over different deadhead coefficients meaningful, the range values are normalized by respective θ values. The values in the next two columns are for minimum liability allocation methods: "MRBR" column represents the percentage of budget recovered when the budget balance property is relaxed and "MRSTMSR" column presents the percent instability value for Method SR. Finally, "INSPR" column represents the percent instability value of the proportional cost allocation scheme, a cost allocation method currently used in practice. The values in this column are obtained by solving a linear program described in Theorem 1.

On the other hand, the first row of Table 1 presents the average values of each performance indicator over all 576 instances. In the next four rows, the average of performance indicators with respect to different deadhead coefficient values are given. Similarly, the average values of the indicators are presented with respect to different number of nodes in rows 6-9, with respect to average number of lanes per node in rows 10-12, with respect to different number of clusters in rows 13-14, with respect to different cluster ratios in rows 15-17 and finally with respect to the supply chain instance indicator in rows 18-19. Finally, the last two rows present the maximum and minimum values of the performance indicators over all the instances generated.

Table 1: Average values of performance indicators of instances with respect to different parameter settings.

		DH %	Range	MRBR	MRSTMSR	INSPR
Average		7.82	1.97	95.66	3.55	8.93
θ	0.25	3.42	1.97	98.12	1.47	3.58
	0.5	6.52	1.97	96.40	2.89	7.15
	0.75	9.37	1.97	94.80	4.26	10.72
	1	11.98	1.97	93.32	5.59	14.29
# of Nodes	50	11.68	1.96	93.75	5.11	13.92
	100	8.30	1.95	95.78	3.43	9.43
	200	6.84	1.98	95.90	3.37	7.60
	500	4.47	1.99	97.21	2.30	4.78
Average # of Lanes per Node	2	8.97	1.94	95.05	3.73	10.16
	5	7.58	1.98	95.69	3.56	8.69
	10	6.92	1.99	96.23	3.37	7.94
# of Clusters	5	7.81	1.97	95.69	3.54	8.90
	10	7.83	1.97	95.63	3.57	8.97
Cluster Ratio	0	8.56	1.98	95.11	4.00	9.99
	0.3	7.84	1.96	95.82	3.39	8.91
	0.6	7.07	1.97	96.05	3.27	7.90
SC	No	4.52	1.98	96.07	3.03	4.85
	Yes	11.12	1.96	95.25	4.08	13.01
Bounds	Max	29.85	2.00	99.43	14.95	42.55
	Min	0.57	1.73	82.82	0.50	0.58

The values in Table 2 presents the averages with respect to the different parameter settings of budget recovery percentage (“PBBR60-PBBR90”) and instability percentage (“PBST90-PBST60”) when the shippers are guaranteed a positive benefit by being allocated at most 60 to 90% of their stand alone costs.

Table 2: Average values of budget recovery and instability percentages with guaranteed positive benefits.

		PBBR60	PBBR70	PBBR80	PBBR90	PBST90	PBST80	PBST70	PBST60
Average		85.90	93.33	97.85	99.39	0.79	2.36	3.83	5.00
θ	0.25	72.44	84.51	96.58	99.30	0.94	INF	INF	INF
	0.5	84.13	95.62	98.18	99.42	0.74	2.39	2.37	INF
	0.75	92.94	96.52	98.31	99.43	0.73	2.37	3.70	2.95
	1	94.10	96.69	98.33	99.43	0.73	2.31	4.53	5.63
# of Nodes	50	82.65	90.79	96.18	98.73	1.82	5.21	7.09	7.38
	100	85.67	93.34	97.93	99.49	0.58	1.91	4.33	6.23
	200	86.73	93.96	98.25	99.55	0.51	1.62	3.30	5.44
	500	88.55	95.25	99.04	99.81	0.23	0.78	1.69	2.99
Average # of Lanes per Node	2	85.16	92.85	97.55	99.32	0.88	2.61	4.49	6.49
	5	86.07	93.46	97.93	99.42	0.74	2.26	3.71	4.99
	10	86.47	93.69	98.07	99.44	0.74	2.20	3.37	3.96
# of Clusters	5	85.92	93.35	97.86	99.38	0.80	2.27	3.61	4.71
	10	85.89	93.32	97.84	99.41	0.77	2.44	4.05	5.27
Cluster Ratio	0	85.30	92.87	97.54	99.28	1.04	2.92	4.04	4.64
	0.3	85.92	93.39	97.91	99.43	0.69	2.24	3.89	5.65
	0.6	86.48	93.74	98.09	99.47	0.63	1.92	3.59	4.66
SC	No	88.19	94.87	98.80	99.68	0.36	1.13	2.30	3.87
	Yes	83.61	91.79	96.90	99.10	1.21	3.60	6.14	8.02
Bounds	Max	99.18	99.66	99.87	99.97	6.55	17.77	35.21	19.56
	Min	67.79	79.09	90.38	96.67	0.03	0.14	0.35	0.89

Our computational experiments show that the ratio of deadhead cost to the total cost is less than 7.9% on average and the maximum value for DH% is approximately 30% and belongs to a supply chain instance. Since there is a lane imbalance in SC instances, it is much more difficult to find a complementary lane to reduce repositioning. The range is approximately 1.97 on average, which is very close to the theoretical bound 2 (although the minimum is 1.73). This shows that, the core does not contain differentiated solutions especially for large instances.

For the game with the minimum liability restriction, on average 95% of the budget is recovered with a stable allocation (column MRBR), which shows that only 45% of the asset repositioning costs are accounted for. This is driven by the supply chain instances and instances with relatively small number of lanes. Also, percent instability value for Method SR (column MRSTMSR) is approximately 3.6% on average. Compared to the theoretical bound, $\frac{\theta}{2} \cong 31.125\%$, this result is substantially better. Although the maximum value of this statistic is relatively large (15%), the minimum value is encouraging (0.5%). The percent instability value for the proportional cost allocation scheme currently used in the industry is 9% on average, with a maximum value of 43% and a minimum value of 0.6%. These results empirically show that Method SR is superior to the proportional cost allocation scheme.

For allocations with positive benefits, approximately 86% of the budget is recovered when the allocated cost of any lane is at most 60% of its stand alone cost (column PBBR60). Note that, when $\theta = 0.25$, $\sigma = 0.6$ imposes the restriction that allocated cost of a lane should be at most 75% of the lane cost. Hence, these values are not surprising. As σ is increased, the percentage of budget recovered also increases, first to 93% then to 98% and finally to 100%. Similarly, the percentage deviation from stability with $\sigma = 0.9$ (column PBST90) is on average 0.79% and slowly increases to approximately 5% as σ decreases. Note that, some of the instances become infeasible with smaller σ values and the infeasible instances are not included in calculating the averages, explaining why the maximum value when $\sigma = 0.7$ is greater than the maximum value when $\sigma = 0.6$.

The effect of different deadhead coefficients is not very significant on the key performance metrics. Note that, the total asset repositioning cost has a linear relationship with the deadhead coefficient. The budget recovery percentage decreases from 98% to 93% due to increase in the total asset repositioning cost. Also, the percent instability values for Method SR (columns MRSTMSR) for the allocation with the minimum liability restriction are approximately 2, 3 and 4 times the values when θ is 0.5, 0.75 and 1 respectively, which empirically suggests a linear relationship with θ as in the upper bound. This result is intuitive, in the sense that, for the allocations with the minimum liability restriction, the cost allocated among the members of the collaboration is the total asset repositioning cost since the shippers are already responsible for their original lane costs. The total asset repositioning cost has a linear relationship with the deadhead coefficient as stated above. This is also true for the proportional cost allocation method as can be seen in Table 1. Finally, percentage of budget recovered when $\sigma = 0.6$ (column PBBR60) increases from 72% to 94% since the constraint on allocating at most 60% of the stand alone cost to the shippers becomes less restrictive as θ increases.

Contrarily, we observe that the number of points is a significant factor on the performance indicators. The ratio of the deadhead cost to the total cost decreases substantially (from 11.7% to 4.5%) as number of points in the network increases. The budget recovery percentage increases from 94% to 97% due to decrease in the total asset repositioning cost. Also, the instability percentage for Method SR decreases from 5.1% to 2.3%. In general, we see that remaining performance measures improve steadily as number of points increases. This phenomenon can be explained as follows; as the number of points in the instances is increased, number of lanes to be covered is also increased in order to keep density of the network constant. Hence, the number of lanes with good synergy increases. Also, as number of points increases, nodes become closer reducing the overall repositioning.

From both Tables 1 and 2, we see that average number of lanes per point is another factor that has a major impact on the results. The ratio of deadhead cost to overall cost decreases from 9% to 7% as average number of lanes per point increases from 2 to 10. Note that this reduction is generally seen in the non-supply chain instances. In the supply chain instances, increasing the average number of lanes does not increase synergy among the lanes due to restricted flow between nodes. In general, all the performance measures improve as number of lanes in the network increases, which suggests that as the collaboration expands the benefits from collaborating increase.

The significance of clusters is assessed by changing the fraction of points within a cluster. Note that, any non-cluster point is distributed uniformly across the total area so it may still fall within a cluster area. The values from both Table 1 and 2 show that increasing the cluster ratio improves the performance measures slightly as opposed to the insignificant affect of the number of clusters.

Finally, we can make the following observations regarding the results on the supply chain instances. The performance metrics for the supply chain instances are significantly worse on average. For example, the deadhead ratio increases to 11.1% from 4.5%. These results are expected since when a supply chain structure is present the requirement for repositioning is increased since a fraction of the points has no incoming arcs (suppliers) and a larger fraction of the points has no outgoing arcs (customers). This shows that repositioning charges increase when the network is mainly composed of vertically integrated supply chain elements, hence inter-firm collaboration becomes beneficial.

As stated before, we have tested the performance of our methods on real-life instances. The smallest of these instances has 15 nodes and 11 edges whereas the largest instance has 784 nodes and 5445 edges and the average number of edges per node is approximately 9.5 over all the instances. The performance indicators of the 18 real-life instances presents similar characteristics as the randomly generated instances, although from the values in Table 3, we see that real-life instances are more challenging due to the imbalance of loads. The interesting observation for these instances is the comparison between the percentage deviation from stability of Method SR and the proportional cost allocation scheme. The values in column INSPR are approximately 3 times of the values in columns MRSTMSR. Furthermore, on average for proportional cost allocation scheme the percent deviation from stability is approximately 25% over all the instances and all deadhead coefficient values which suggests that there exists a significant risk for the disintegration of the collaboration. For example, when $\theta = 1$, the deviation from stability is approximately 40%, which might be too high of a value for depending on notions like "cost of collaborating" or "limited rationality" to keep the collaboration together.

Table 3: Average values of performance indicators for real-life instances.

		DH %	Range	MRBR	MRSTMSR	INSPR
Average		21.68	1.84	89.46	8.31	24.54
	0.25	10.55	1.84	95.06	3.48	10.07
θ	0.5	18.92	1.84	90.93	6.79	19.86
	0.75	25.76	1.84	87.43	9.96	29.42
	1	31.49	1.84	84.42	13.00	38.81

2.8 Conclusions

In this chapter, we consider the cost allocation problem of a collaborative transportation procurement network and design effective and computationally efficient cost allocation methods to assist shippers to manage their collaboration structures. A good cost allocation mechanism should attract shippers to the collaboration, enable easier contractual agreements, and help to maintain the collaboration together. Although according to the CEO of Nistevo Network, Kevin Lynch, *"The key to understanding* *collaborative logistics lies in recognizing how costs are distributed in a logistics network,"* current practice only employs simple allocation mechanisms that may be very instable.

Our main concern is to develop implementable mechanisms to allocate gains from the collaboration in a stable manner to ensure the continuity of the collaboration. Due to several different challenges faced in establishing and managing a shippers' logistics network, we identify several desirable cost allocation properties. However, we also show that no method can ensure allocations with all of these properties. As a result, we design several algorithms that generate allocations with worst case bounds on the relaxed properties.

CHAPTER III

COLLABORATION FOR TRUCKLOAD CARRIERS

3.1 Introduction

Trucking is the backbone of U.S. freight movement. According to the American Trucking Association (ATA), the trucking industry's share of the total volume of freight transported in the United States was 68.9% in 2005. The US trucking industry produced an annual revenue of $623 billion by hauling 10.7 billion tons in 2005, which was 84.3% of the nations freight bill. Mostly because of historically high fuel prices, the trucking industry's operating expenses are higher than ever. "Each penny increase in diesel costs the trucking industry $381 million over a full year" [3]. Eventually, these increased operating costs are reflected in the prices charged to the shippers.

We focus on the full truckload segment of the trucking industry. Razor-thin profit margins have forced full truckload carriers to search for ways to cut costs. One way to do so is through mergers and acquisitions. There are economies of scale, in the form of increased buying power and reduced marketing and administrative expenses, and economies of scope, in the form of reduced repositioning costs. When shipment density increases, due to a merger or acquisition, the likelihood of geographical synergies between shipments increases, which in turn will reduce costly empty travel between consecutive shipments. However, mergers and acquisitions may be beyond reach in many situations. In that case, companies may consider collaboration as a means to achieve some economies of scale and scope.

Several trucking companies already sought out collaboration to improve their overall performance. For example, six of the largest truckload carriers in the United

States, namely J.B. Hunt Transport Services, Werner Enterprises, Swift Transportation, M.S. Carriers, U.S. Xpress Enterprises and Covenant Transport, agreed on the partnership under a collaborative logistics network called Transplace.com. This partnership allows the trucking companies and companies requesting transportation services to meet and match their hauling capacity and transportation demand using an internet based management technology [46]. The collaborative logistics network attempts to find the best assignment of shipment requests to the carriers based on the operating costs of the carriers. The carriers benefit from this collaborative approach through better truck utilization and lower driver turnover.

Collaboration among full truckload carriers may result in significant cost savings. However, realizing the full potential of such a collaboration is challenging. Ideally, a centralized decision maker with complete information about all participating carriers determines the optimal assignment of shipment requests to the carriers and identifies the minimum cost routes for the carriers. Besides the fact that a complex optimization problem needs to be solved, companies may not be willing to share the necessary information. Trust is a central issue for selfish individuals in collaborative relationships, especially in horizontal collaborations where competing entities collaborate. Another complexity that may arise is that in a centralized solution the cost for one or more of the carriers may increase, in which case these carrier will not agree to the collaboration unless they are compensated with side payments.

In case centralized decision making is not viable, alternative collaboration mechanisms need to be explored. That is the focus of our work. We design and analyze relatively simple mechanisms that allow full truckload carriers to initiate and manage a collaboration. Because the carriers are guided by their own self-interests, any proposed mechanism to manage the collaboration's activities has to yield collectively and individually desirable solutions. Also, because of the trust issues associated with horizontal collaborations, the proposed mechanisms cannot rely on the availability of full information. The challenge is to design mechanisms that are simple, fast, implementable, yet effective in terms of benefits to all participants. We will show that bilateral lane exchanges, where a lane represents a full truckload shipment request from an origin to a destination, satisfy all these requirements. We will also show that with more information sharing and with side payments, almost all the potential cost savings can be realized.

The remainder of the chapter is organized as follows. In Section 3.2, we illustrate economies of scope in truckload transportation and discuss how collaboration may exploit economies of scope. In Section 3.3, we briefly review related work. In Section 3.4, we introduce and analyze the multi-carrier lane covering problem, which is the optimization problem that needs to be solved to determine the maximum benefits of collaborating. In Section 3.5, we present and investigate lane exchange mechanisms for different collaborative environments in terms of information sharing and side payment options.

3.2 Economies of Scope - An Example

We demonstrate the potential benefits of carrier collaboration with a simple example. Consider a network with four cities and two carriers A and B. We assume that the cost of traveling between two cities is the same for both carriers and, for simplicity, that there is no difference in cost between traveling loaded or empty. We further assume that Carrier A has a contract in place to serve lane (3,2) and that Carrier B has a contract in place to serve lane (1,2). Figure 4 shows the relevant information.

Suppose that new shipment requests for lanes (2,4), (4,1), and (2,3) arrive sequentially over time and that they are assigned to the carriers based on a marginal cost comparison at the time of the arrival. Hence, Carrier A gets lane (2,4) (marginal costs are 1 and 1.73, respectively), Carrier B gets lane (4,1) (marginal costs are 2.73 and 1.73, respectively) and lane (2,3) (marginal costs are 1 and 1, respectively). The

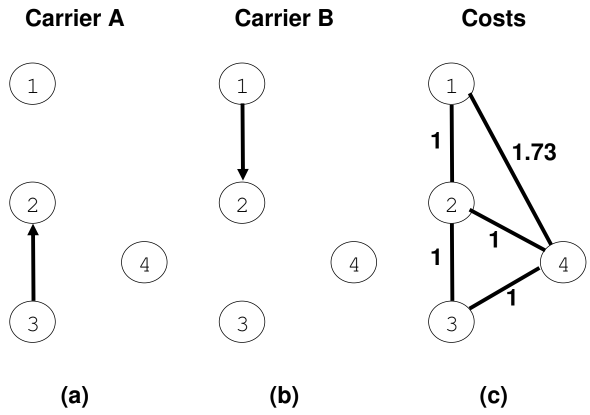

Figure 4: Cost information and existing contracts.

minimum cost routes covering the lanes that have to be served for both carriers are presented in Figure 5, where a dashed line represents repositioning (or empty travel). The corresponding transportation costs are 3 for Carrier A and 4.73 for Carrier B.

If Carrier A and Carrier B collaborate and exchange lanes (2,4) and (2,3), they significantly reduce their repositioning costs and the transportation costs become 2 for Carrier A and 3.73 for Carrier B. In fact, this assignment of lanes to carriers and the accompanying routes is the best possible and would have been determined by a central decision maker.

Finally, note that the setting considered above does not consider any characteristics related to time, e.g., no travel time and dispatch and/or delivery windows. Incorporating time related characteristics will make the analysis far more complicated.

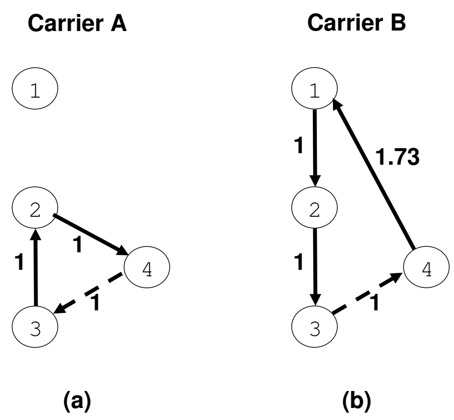

Figure 5: Minimum cost routes.

3.3 Literature Review

In this section, we briefly review related work in the literature. The literature on truckload transportation procurement can be roughly divided into two streams: centralized and auction-based approaches.

Much of the research on centralized approaches considers only a single-carrier setting and focuses on optimizing the routing decisions for that carrier, i.e., finding continuous-move paths and tours covering the lanes and minimizing repositioning costs. Ergun et al. [18] introduce the Lane Covering Problem (LCP), i.e., the problem of finding a minimum-cost set of cycles covering a given set of lanes, and show that LCP is polynomially solvable. Ergun et al. [18] and Ergun et al. [17] consider variants of LCP that include practically relevant side constraints, such as dispatch windows and driver restrictions, and show that these variants are NP-hard. Their proposed heuristics appear to be effective. Moore et al. [28] develop a mixed integer programming model and a simulation tool to facilitate the centralized management of interstate truckload shipments, which involves selecting carriers and dispatching shipments.

The research on truckload transportation procurement using auction-based methods mostly focuses on the assignment of lanes from a single shipper to a set of carriers so as to minimize the total transportation cost. Combinatorial auctions are most suitable for transportation procurement auctions, because they allow the capturing synergies between shipment requests through bundle bids. The shipper has to solve a winner determination problem to assign lanes to the carriers. This involves solving a "set packing" type problem, which is know to be NP-hard. Sheffi [39] provides a survey on combinatorial auctions in procurement of truckload transportation services. He observes that solving large instances is quite challenging: "Real problems, which may include thousands of lanes, dozens or hundreds of carriers, and millions of combination bids, are more challenging and require special code." See also De Vries and Vohra [11], Caplice and Sheffi [9], and Song and Regan [41]. Elmaghraby and Keskinocak [13] give an overview of the challenges in designing and implementing combinatorial auctions and presents a case study on how Home Depot utilizes combinatorial auctions in the procurement of transportation services. An et al. [2] investigate how the participants in a combinatorial auction should select their bids as evaluating the bids for all possible bundles is not practical for both the bidders and the auctioneer. Ledyard et al. [27] discuss the implementation of "combined-value auctions" for the purchasing of transportation services by Sears Logistics Services.

Our perspective in this chapter is quite different, we are not interested in the transportation procurement process per se, i.e., in how shippers select carriers and assign lanes to carriers, but instead on how carriers can reduce cost by collaborating given the set of lanes that they have to serve. This is especially relevant in a more dynamic setting in which the set of lanes that a carrier has to serve changes frequently over time. The goal is to design simple lane exchange processes to facilitate collaboration among carriers.

3.4 The Multi-Carrier Lane Covering Problem

To assess the potential benefit of collaborating, the optimal assignment of lanes to carriers has to be determined and the optimal set of cycles covering the lanes assigned to each carrier needs to be found. We refer to this optimization problem as the Multi-Carrier Lane Covering Problem (MCLCP).

The total transportation costs breaks down into two components: lane covering costs and repositioning costs. The cost of covering a lane depends only on the carrier assigned to execute the associated shipment. Therefore, if the carriers have identical cost structures, the lane covering costs are independent of the assignment and routing decisions. On the other hand, the repositioning costs depend on the synergy among (or complementarity of) the set of lanes assigned to a carrier (irrespective of any differences in carrier cost structures), and thus on the assignment and routing decisions.

We assume that our starting point is a set of carriers each with two sets of lanes: lanes that have to be executed by the carrier and lanes that can be executed by any carrier. (The set of lanes that have to be executed by the carrier may be empty.) Of course only the lanes that can be executed by any carrier can be re-assigned during the optimization. We further assume that the cost of traversing an arc with a full truckload is independent of the direction along the arc and that the cost of repositioning along an arc is a percentage of the full truckload cost. These costs can be different for the different carriers. There are no capacity restrictions; each carrier is able to handle all the assigned lanes. Finally, we observe that we ignore some relevant practical considerations, such as pickup and delivery windows, driver restrictions, etc.

Thus, MCLCP is defined on a complete graph $G = (N, A)$, where N is the set of nodes $\{1, \ldots, n\}$ and A is the set of arcs. The set of lanes that can be covered by any carrier is denoted by $L \subseteq A$ whereas the set of lanes that can only be covered by

Carrier k is denoted by $I^k \subseteq A$. Let K be the set of carriers offering transportation services and let c_{ij}^k denote the cost of traversing arc (i,j) with a full truckload for Carrier k ($k \in K$). The repositioning cost coefficient of Carrier k is denoted by θ^k ($K \in K$), where $0 < \theta^k \leq 1$, and hence the repositioning cost along an arc $(i,j) \in A$ is equal to $\theta^k c_{ij}^k$ if traversed by Carrier k. The objective is to cover the lanes at minimum cost.

MCLCP can be formulated as an integer linear program as follows:

$$z_L = \min \sum_{k \in K} \sum_{(i,j) \in L} c_{ij}^k x_{ij}^k + \sum_{k \in K} \theta^k \sum_{(i,j) \in A} c_{ij}^k z_{ij}^k \tag{39}$$

$$s.t. \sum_{j \in N} x_{ij}^k - \sum_{j \in N} x_{ji}^k + \sum_{j \in N} z_{ij}^k - \sum_{j \in N} z_{ji}^k = 0 \quad \forall i \in N,\ \forall k \in K \tag{40}$$

$$\sum_{k \in K} x_{ij}^k \geq r_{ij} + \sum_{k \in K} I_{ij}^k \quad \forall (i,j) \in L \tag{41}$$

$$x_{ij}^k \geq I_{ij}^k \quad \forall (i,j) \in I^k,\ \forall k \in K \tag{42}$$

$$z_{ij}^k \geq 0 \quad \forall (i,j) \in A,\ \forall k \in K \tag{43}$$

$$x_{ij}^k, z_{ij}^k \in \mathbb{Z}. \tag{44}$$

Variable x_{ij}^k represents the number of times lane $(i,j) \in L$ is traversed with a full truckload by Carrier k and variable z_{ij}^k represents the number of times arc $(i,j) \in A$ is traversed empty (i.e., to reposition) by Carrier k. Furthermore, I_{ij}^k represents the number of times lane $(i,j) \in L$ has to be traversed with a full truckload by Carrier k and r_{ij} represents be the number of times lane $(i,j) \in L$ has to be traversed with a full truckload by any of the carriers. Let r be the matrix of r_{ij}'s. The objective is to minimize the sum of the lane covering costs of all the carriers plus the sum of the repositioning costs of all the carriers. Constraints (40) are flow balance constraints for all the nodes in the network and all the carriers offering service. Constraints (41) ensure that the transportation requirement of each lane is satisfied. Constraints (42) ensure that carrier specific requirements are satisfied. A solution to MCLCP gives the optimal assignment of the lanes to the carriers and the optimal routes for covering the lanes assigned to each carrier. The objective function represents the minimum cost for serving all shipment requests.

Next, we examine the complexity of MCLCP. There are a few characteristics that impact the complexity: the number of carriers, whether or not carriers have sets of lanes that they have to serve, and whether or not the carriers have identical cost structures. We have the following set of results.

Lemma 6 *When carriers have no sets of lanes that only they can serve and when their cost structures are identical, then MCLCP can be solved in polynomial time for any number of carriers.*

Proof: Follows from the fact the that the unconstrained version of the single-carrier lane covering problem is polynomially solvable [18]. □

Theorem 3 *When there are three or more carriers with non-identical cost structures, then MCLCP is NP-hard.*

Proof: See Appendix A. □

As MCLCP is NP-hard, the LP-relaxation may not yield an integral solution. Next, we present some observations concerning the integrality gap.

Lemma 7 *The integrality gap for MCLCP is less than or equal to* 100%*.*

Proof: Let z^*_{LP} be the objective function value of the LP-relaxation, then we have

$$z^*_{LP} \geq \sum_{(i,j)\in L} \min_{k\in K} c^k_{ij} r_{ij} + \sum_{k\in K} \sum_{(i,j)\in L} c^k_{ij} I^k_{ij},$$

because the right hand side represents the cost of traversing each lane in L using a minimum cost option. Furthermore, traversing each lane in L with its lowest cost option and returning back empty along that lane is an integral feasible solution to

MCLCP. Because $\theta^k \leq 1 \quad \forall k \in K$, the objective function value of this feasible solution, say z_{IP}, satisfies

$$z_{IP} \leq 2\,\{ \sum_{(i,j)\in L} \min_{k\in K} c_{ij}^k r_{ij} + \sum_{k\in K} \sum_{(i,j)\in L} c_{ij}^k I_{ij}^k \}.$$

Hence, the integrality gap is less than or equal to 100%. □

Next, we present an instance with three carriers for which the integrality gap is 33% (see Figure 6). The carriers do not have lanes that have to be served by them. The bold black arcs represent the lanes in L. For each carrier, the cost of traversing a lane is given next to the arc. (Note that the costs for each carrier satisfy the triangular inequality.)

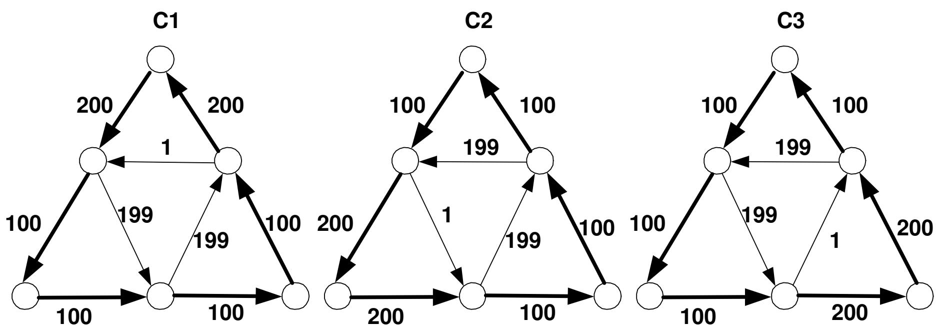

Figure 6: An adversary instance.

The objective function value of the LP-relaxation is 601.5 and the corresponding flows for each carrier are shown in Figure 7. The optimal objective function value of the IP is 800, which gives an integrality gap of $\frac{198.5}{601.5} = 33\%$. By slightly modifying the costs, we can make the integrality gap $\simeq 33.3\%$.

Next, we discuss exact and heuristic algorithms for solving MCLCP. We can, of course, simply use a commercially available integer programming solver and the formulation presented earlier. Furthermore, if we are willing to accept solutions that are not necessarily optimal, but are guaranteed to have a value within a given percentage of the optimal value, we can specify a relative optimality stopping criterion

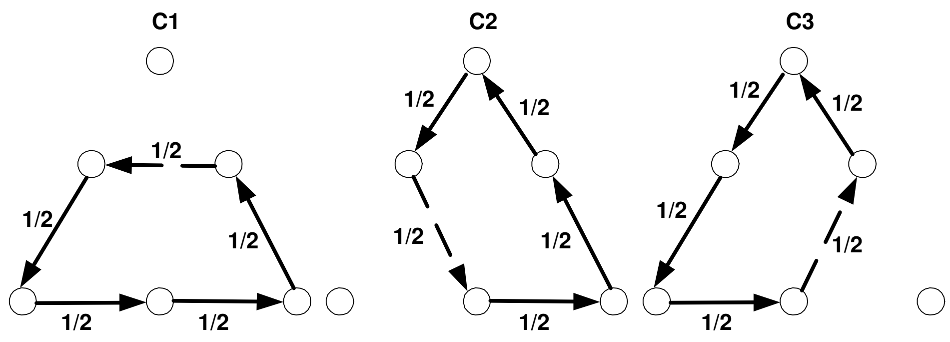

Figure 7: Optimal solution to the LP-relaxation.

and solve the integer program with relaxed precision. This may significantly reduce the computation time, because a significant amount of time may be spent on proving optimality.

To reduce the computation time further, we can exploit solution properties. Any feasible solution is composed of *cycles* covering lanes assigned to a particular carrier. Consider Figure 8, which shows the routes assigned to a single carrier. The fractional solution in Figure 8(a) consists of the two cycles shown in Figure 8(b). The *flow* on arc (i, j) for carrier k is equal to the value of variable x_{ij}^k. This flow is composed of an *integral flow*, the integer part, and a *fractional flow*, the remaining part. For example, the flow on arc (1,2) is equal to 2.6 and hence, the integral and fractional flows are 2 and 0.6, respectively. We define the *flow on a cycle* as the minimum flow on any of the arcs of the cycle. For example, the flow on Cycle 1 in Figure 8(b) is equal to $\min\{2.6, 4, 4, 2.6, 2.6\} = 2.6$. Any integral feasible solution to MCLCP can be decomposed into cycles with integral flows. Every decomposition of a fractional solution to MCLCP, such as the solution to the LP-relaxation, has at least one cycle with a fractional flow. Note that this does not necessarily require all x_{ij}^k values to be fractional along the cycle. For example, arcs $(2, 3)$ and $(3, 4)$ of Cycle 1 have integral flows. If there is a variable x_{ij}^k that is fractional, then, because of the flow balance constraints, there exists at least two arcs, one preceding (i, j) and one succeeding

(i,j), with fractional flow on them. Therefore, there exists at least one *loop*, i.e., an undirected cycle, containing arc (i,j) with fractional flows on all the arcs in that loop. Because arc $(1,2)$ has a fractional flow of 0.6, there exists a loop containing arc $(1,2)$. The loop is shown in Figure 8(c); the fractional flow on the arcs $(4,5)$ and $(5,1)$ is equal to 0.6, whereas the fractional flow on the arcs $(4,7)$, $(7,6)$, and $(6,2)$ is equal to 1 - 0.6 = 0.4.

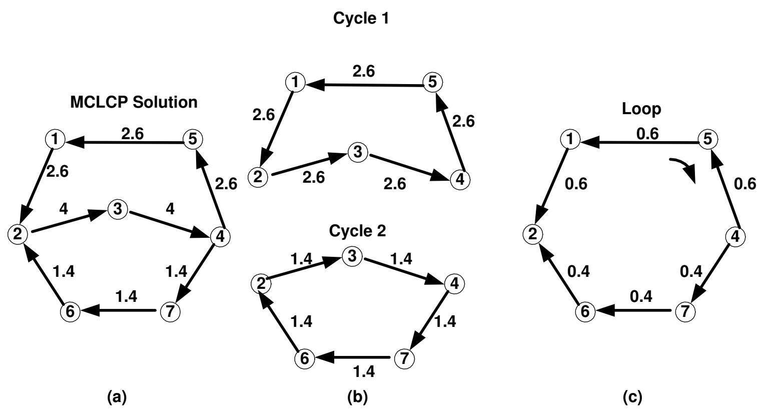

Figure 8: A solution and corresponding cycles and loop.

One way to exploit these properties, is to fix and remove integral flows from the solution to the LP-relaxation and solve the resulting integer program. As the size of the integer program may be a lot smaller, it may be solved more quickly.

We have carried out a computational study to evaluate the performance of the methods discussed above on randomly generated instances. The instances are generated as follows: All locations are in a 1,000 $\times$ 1,000 mile square region. Two parameters control the size of the instance: the number of locations and the average number of lanes incident to a location (which, of course, controls the total number of lanes). Two further parameters control the spatial characteristics of the instance: the number of clusters and the ratio of cluster points to total points. The spatial control parameters allow us to introduce dense regions with many origin and destination locations, representing urban areas, and to vary the ratio between urban and rural locations. Clusters, or urban areas, are uniformly distributed in the square and each location in a cluster is generated using a standard Normal distribution. The remaining, rural, locations are distributed uniformly across the square. Lanes are created by randomly picking an origin and a destination, where we ensure that no lanes are generated between two locations in the same cluster. The cost of traveling between locations with a full truckload is equal to the Euclidean distance between the locations and the repositioning cost coefficient for all three carriers is equal to 0.75.

We generated 270 instances with 100, 200, and 300 locations, with 2, 4, and 6 as the average number of lanes incident to a location, with 4 clusters, and with 45% of the locations within clusters. There are three carriers in all instances. To introduce lanes that have to be served by a carrier, we add lanes for each carrier with 0.25, 0.5, and 1 as the average number of lanes incident to a location, respectively. We want carrier cost structures to be similar but not identical, as this is what typically happens in practice. For each location, we generate two dummy locations within a small radius of the original location (using a standard Normal distribution), giving us three similar but not identical networks. We take the Euclidean distance between the locations as the cost of traveling between locations with a full truckload thus giving us three similar but not identical cost structures.

To analyze the performance of the different solution approaches, we study the optimality gap. The results are found in Table 4, where we present the number of instances for which the LP-relaxation yielded an integral solution (#int), the average integrality gap as a percentage of the value of the integral solution (avg. gap), the maximum encountered integrality gap (max. gap), the average and maximum optimality gap when we solve instances with relaxed precision (avg. gap RP and max. gap RP) and the average and the maximum optimality gap of the cycle fixing heuristic (avg. gap CFH and max. gap CFH).

Table 4: Integrality gap of the LP relaxation and optimality gap values of the proposed algorithms.

# nodes	# int.	avg. gap	max. gap	avg. gap RP	max. gap RP	avg. gap CFH	max. gap CFH
100	22	0.0014	0.0115	0.0008	0.0324	0.0000	0.5177
200	5	0.0014	0.0057	0.0010	0.0119	0.1191	0.3058
300	2	0.0014	0.0046	0.0013	0.0063	0.0839	0.2283

The associated average and maximum solutions times are given in Table 5.

Table 5: Computational times of the proposed algorithms in CPU seconds.

# nodes	avg. time IP	max. time IP	avg. time RP	max. time RP	avg. time CFH	max. time CFH
100	59	231	50	142	55	216
200	1128	5064	780	2077	1620	8275
300	9822	83514	3201	8356	3165	6672

The computational experiments demonstrate that integer programming with relaxed precision and the cycle fixing heuristic are effective heuristics for solving MCLCP as the maximum optimality gap over all instances are less than 0.04% and 0.6%, respectively. We further observe that the computational advantages of the heuristics become apparent only for instances with 300 nodes. The reason for this is that the solution of a number of LP-relaxations does not payoff until the IP search trees get sufficiently large. It is already visible that as the size of the instances increases, the cycle fixing heuristic will become the fastest because also the search tree for the integer program with relaxed precision continue to grow.

3.5 Carrier Collaboration

We consider settings in which several carriers collaborate by means of bilateral lane exchanges, i.e., by two carriers exchanging shipment requests, sometimes accompanied by side payments. The goal is to reduce any geographic imbalances in the carriers' networks so as to reduce repositing costs. The carriers are selfish in the sense that their sole objective is to minimize their own costs, although they may be willing to share a portion of the benefits of a lane exchange if they believe this is required for the exchange to happen. The carriers have agreed up front on the rules and regulations governing the lane exchange process, primarily the level of information sharing and whether or not side payments can be offered. We have chosen to focus on bilateral lane exchanges because they represent the simplest setting for exchanges and thus facilitate the analysis. However, the ideas discussed can be relatively easily extended to lane exchanges involving more than two carriers. Lane exchanges with multiple carriers do significantly increase the computational challenges. Furthermore, our computational experiments show that most of the benefits of collaboration can be achieved through bilateral lane exchanges.

Given a lane exchange mechanism, i.e., a set of rules and regulations governing the lane exchange process, a carrier determines his own strategy. That is, a carrier decides himself on the lane to offer to the other carrier and on the side payment, if allowed by the exchange mechanism. Depending, again, on the exchange mechanism, a carrier may have only his own individual cost and network information or may have information about the other carriers' costs and networks. We consider only bilateral lane exchange mechanisms, i.e., any lane exchange occurs between two carriers. Of course, this does not prevent several carriers from collaborating. It just means they do so through a sequence of bilateral lane exchanges. In order for a lane exchange to happen, both carriers must agree to the exchange, hence the result of the exchange must be acceptable to both carriers. This property ensures that the carriers are always better off participating in the collaboration, which is in line with the selfishness of the carriers.

The level of information sharing between carriers and the decision to allow or disallow side payments are the two primary factors defining a lane exchange mechanism. Carriers may be hesitant to share information with their competitors and therefore may decide not to do so. This, of course, forces them to select the lane to offer and to decide on the side payment based solely on their own cost and network information, so somewhat blindfolded; some beneficial lane exchange opportunities will not be recognized or identified. Limited information sharing not only limits the number of possible exchanges considered but also decreases the possibility of an exchange being accepted, because a carrier is more likely to offer a lane or a side payment that is unacceptable to the other carrier. With information sharing, carriers can estimate the counter strategies of their collaborators, which makes it easier to identify the best overall strategy. Hence, increasing the level of information sharing usually increases the value of the collaboration. However, as will be shown later, this is not always the case. Side payments, or compensations, allow carriers to share some of their benefits from an exchange with the other carrier. Because an exchange will only take place if both carriers agree to it, a carrier has to make his offer attractive to the other carrier so as to make it acceptable. Suppose that Carrier A offers a lane with low synergies with the other lanes in his network. Due to the low synergies, the operating cost for this lane is high and if the offer is accepted, it will likely result in considerable cost savings. Although Carrier B may be able to handle this lane at a lower cost, because of possible synergies with the other lanes in his network, the exchange may be unacceptable for Carrier B as his costs may still increase. In this case, Carrier A can share some of his cost savings with Carrier B via a side payment and make the exchange offer acceptable to Carrier B. Hence, allowing side payments increases the number of acceptable exchanges and so is usually beneficial. Again, as will be shown later, this is not always the case.

Designing an effective lane exchange mechanism presents several challenges. The mechanism should identify exchanges that are acceptable to the individual carriers while trying to reach a system optimal solution (i.e., an optimal solution to the MCLCP). Furthermore, the mechanism should be computationally viable and be able to determine a candidate exchange without having to enumerate all possible exchanges.

Different levels of information sharing and whether or not side payments are allowed present different challenges, all of which a mechanism should accommodate. For example, without information sharing and side payments, the carriers are myopic when deciding on a lane to offer and have no means to make this offer acceptable to the other carrier. If side payments are allowed, determining a side payment for the offered lane is non-trivial because of the lack of information about the other carrier. With full information sharing, the number of alternative strategies is quite large, which makes the exchange process complicated, especially if side payments are allowed. A carrier has to evaluate all lanes and simultaneously evaluate all possible side payments, since the benefits of the a lane exchange depend on both the lane and the associated side payment.

3.5.1 Lane Exchanges - Illustrative Examples

We demonstrate the challenges associated with designing lane exchange mechanisms by means of examples. We consider a network with five cities and three carriers, namely A, B and C. The cost of covering a lane is assumed to be the same for all carriers; the costs are shown in Figure 9(a). For simplicity, we also assume that the cost of traversing an arc with an empty truck is equal to the cost of traversing an arc with a loaded truck. The lanes of each carrier and the optimal way to cover them are shown in Figure 9(b), (c), and (d), respectively, where a dashed line represents repositioning. The corresponding transportation costs are 3, 4.73, and 5.46 respectively.

Consider a lane exchange setting in which there is no information sharing and there are no side payments. In this setting, the only task of a carrier is to select the lane to offer to the other carrier. Suppose that Carrier A and B try to exchange lanes. Because a carrier has no information about the other carrier, he is likely to choose a lane that is costly to handle, such as the highest marginal cost lane or the highest per mile marginal cost lane, hoping that the lane has a greater synergy within the other carrier's network. Hence, Carrier A may offer either lane (3,2) or lane (2,4); they are equivalent from Carrier A's point of view. Carrier B may offer lane (2,3) since it has the highest marginal cost (and the highest per mile marginal cost). As a consequence, if Carrier A decides to offer lane (3,2), the exchange will not be acceptable, but if Carrier A decides to offer lane (2,4) instead, both carriers will agree on the exchange as it reduces their costs to 2 and 3.73, respectively.

Allowing side payments results in more acceptable lane exchanges. Suppose that Carrier B and C try to exchange lanes. Furthermore, let Carrier B offer lane (2,3) as it is the highest marginal cost lane and, similarly, let Carrier C offer lane (1,4).

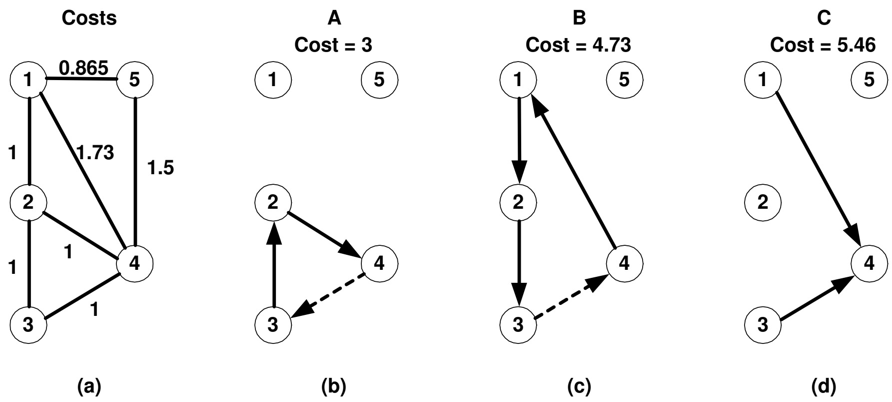

Figure 9: Collaboration among three carriers.

If the carriers exchange these lanes, then their costs become 5.46 and 3, respectively. Hence, Carrier B will not agree to the exchange. However, if Carrier C is willing to share some of his benefits of 5.46 - 3 = 2.46 and offer a side payment of, say one unit, then Carrier B will accept the improved offer as it results in a cost of 5.46 - 1 = 4.46, which is less than 4.73. At the same time, Carrier A still benefits too as his costs are 3 + 1 = 4, which is still less than 5.46. This shows that carriers can benefit from allowing side payments. However, it is not obvious how a carrier should determine the amount of the side payment, especially without information about the other carrier's network.

Information sharing allows carriers to calculate the value and outcome of any possible exchange. Therefore, carriers can initiate lane exchanges with higher benefits and higher likelihood of acceptance. Suppose that Carrier A and B try to exchange lanes and share network and cost information. Carrier B realizes that if he offers any lane but (2,3) Carrier A will reject the exchange, regardless of the lane offer by Carrier A itself. Therefore, Carrier B offers lane (2,3). Similarly, Carrier A realizes that Carrier B will offer lane (2,4) and so he chooses to offer lane (2,4). The exchange will be accepted as both carriers reduce their costs, to 2 and 3.73 respectively.

Not surprisingly, with information sharing determining a strategy becomes considerably more involved. Evaluating each possible lane exchange is not only challenging and time consuming, but may not yield a dominant solution, i.e., a unique lane exchange that results in the largest cost savings for both carriers. Consider a network with eight cities and two carriers A and B. The cost of covering a lane is assumed to be the same for all carriers; the costs are shown in Figure 10(a). Furthermore, we assume that the cost of traversing an arc with an empty truck is equal to the cost of traversing an arc with a loaded truck. The networks of Carrier A and B are shown in Figure 10(b). The carriers share network and cost information. If Carrier A offers lane (2,3), then his cost savings will be 0 in case Carrier B offers (3,2) or 2 if Carrier B offers (7,6). On the other hand, if Carrier A offers lane (6,7), then his cost savings will be 2 in case Carrier B offers (3,2) or 0 if Carrier B offers (7,6). The situation is identical for Carrier B. That is, there is no dominant solution and the carriers are still uncertain about their strategies even though they have perfect information. The reason for the uncertainty is that the carriers do not know the strategy of the other carrier.

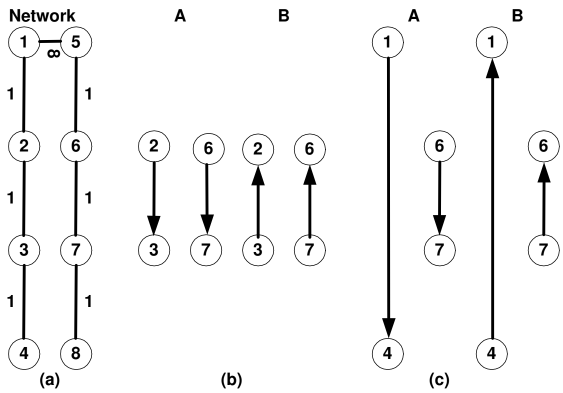

Figure 10: Collaboration with information sharing.

Another phenomenon is that information sharing may not lead to the best possible lane exchange. Suppose the networks of Carrier A and B are as shown in Figure 10(c).

Both Carrier A and B will try to get rid of their longest lane and offer (1,4) and (4,1) respectively. This lane exchange does not lead to any cost savings, so the exchange will not be accepted. However, there exist lane exchanges resulting in cost savings for both carriers, e.g., Carrier A offering (1,4) and Carrier B offering (7,6).

Finally, although information sharing is generally helpful, it is easy to show that this is not always the case. Suppose the networks of Carrier A and B are as shown in Figure 11. Suppose that Carrier A can offer only lanes (4,1) and (8,6) and Carrier B can offer only lanes (5,3) and (10,7). Without the information sharing, the carriers will likely offer lanes with the highest marginal costs, i.e., (4,1) and (10,7) respectively, resulting in cost savings of 1 for both carriers. With information sharing, if Carrier A offers (4,1), then his cost savings will be 1 in case Carrier B offers (10,7) or 0 if Carrier B offers (5,3). On the other hand, if Carrier A offers lane (8,6), then his cost savings will be 2 in case Carrier B offers (10,7) or 0 if Carrier B offers (5,3). As the potential payoffs from offering lane (8,6) dominate the alternative, Carrier A chooses to offer lane (8,6). Similarly, Carrier B chooses to offer lane (5,3). Thus their cost savings will be 0 whereas without information sharing their cost savings would have been 1.

Also with information sharing, allowing side payments is generally beneficial as it increases the number of acceptable lane exchanges. However, determining an appropriate side payment with an offered lane is even more difficult in this setting. The cost savings for a carrier depend on the lane offered by the other carrier, so even if a carrier is willing to share the benefits, he is uncertain about the actual cost savings of a lane exchange until the moment of the exchange. Given a candidate lane to offer calculating an appropriate side payment for all possible counter offers may be too time consuming. Determining an appropriate side payment is further complicated by the fact that the other carrier may also offer a side payment.

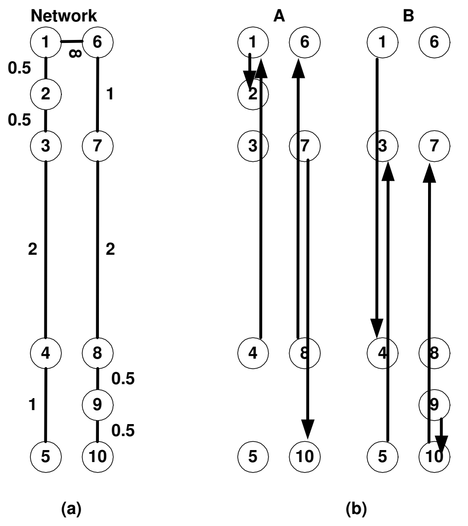

Figure 11: An adversary example for information sharing.

3.5.2 Lane Exchange Mechanisms

We propose and analyze lane exchange mechanism for four different carrier collaboration settings: no information sharing and no side payments (NINS), no information sharing with side payments (NIWS), information sharing without side payments (WINS), and information sharing with side payments (WIWS). We first discuss some common properties of these lane exchange mechanisms and then describe them in detail and discuss their advantages and disadvantages.

Let Carrier A and B be the two carriers in a lane exchange process. Let L^A and L^B be the sets of lanes the carriers can offer for exchange. (Recall that some lanes may have to be served by the carrier itself.) Let c_{ij}^A represent the cost of covering arc (i,j) for Carrier A, i.e., the cost of traversing the lane with a full truckload, and let θc_{ij}^A be the repositioning cost for Carrier A along that same arc. Let $z_A^*(L^A)$ be the minimum cost of covering all the lanes in the lane set L^A by Carrier A, i.e., the optimal objective function value of the lane covering problem (LCP) for Carrier A. The marginal cost $MC_{ij}^A(L^A)$ of lane (i,j) for Carrier A is $z_A^*(L^A) - z_A^*(L^A \backslash (i,j))$. Furthermore, let p_{ij}^A denote the offer price for lane (i,j) by Carrier A, i.e., the amount that Carrier A will pay to Carrier B if the lane exchange is accepted. This offer price includes any side payment by Carrier A. The "payoff" of an exchange to a carrier refers to the cost savings for the carrier if the lane exchange is accepted and performed. The payoff to Carrier A is denoted by $\pi^A_{ij,uv}$, where (i,j) represents the lane offered by Carrier A and (u,v) the lane offered by Carrier B, and similarly the payoff to Carrier B is denoted by $\pi^B_{uv,ij}$. The payoffs are equal to

$$\pi^A_{ij,uv} = \begin{cases} MC^A_{ij}(L^A) - MC^A_{uv}(L^A \backslash (i,j) \cup (u,v)) - p^A_{ij} + p^B_{uv} & \text{if exchange occurs,} \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\pi^B_{uv,ij} = \begin{cases} MC^B_{uv}(L^B) - MC^B_{ij}(L^B \backslash (u,v) \cup (i,j)) - p^B_{uv} + p^A_{ij} & \text{if exchange occurs,} \\ 0 & \text{otherwise.} \end{cases}$$

Note that the payoffs depend on both lanes. That is, the marginal cost of an offered lane depends on the lane that the other carrier offers. Therefore, strategies that ignore the other carrier's actions are less likely to be effective.

3.5.2.1 No Information Sharing - No Side Payments

In this most basic lane exchange setting, a carrier only needs to select the lane to offer to the other carrier. In the lane exchange mechanism we propose, the carriers offer the lanes with the highest marginal costs. That is, Carrier A offers $argmin_{(i,j)\in L^A}\{MC^A_{ij}(L^A)\}$ and Carrier B offers $argmin_{(u,v)\in L^B}\{MC^B_{uv}(L^B)\}$. The selection strategy is completely based on a carrier's individual network and ignores the other carrier's network. The motivation for this selection strategy is that it maximizes the known benefit from a potential exchange as it discards the lane with highest cost.

An alternative selection strategy is to offer the lane with the highest marginal cost per mile, which is an indicator for the repositing requirement of the lane. The motivation for this selection strategy is that the highest marginal cost lane may be a long lane rather than a lane with low synergies, and therefore it may not be attractive to the other carrier. On the other hand, the lane with the highest marginal cost per mile has limited or no synergies, and therefore may be more likely to have synergies with lanes of the other carrier. As such, this selection strategy attempts to maximize the possibility of acceptance of the lane exchange.

The computational requirements of both selection strategies are small as the exchange mechanism is simple and only requires the ranking of the lanes according to their marginal costs (or per mile marginal costs).

The following algorithm evaluates the benefits of this exchange mechanism for two carriers A and B, where we assume that the carriers continue to exchange lanes as long as beneficial lane exchanges are identified:

NINS:

Step 1: Rank the lanes that can be offered according to marginal costs (or per mile marginal costs) for both carriers and select the best one from each to form the lane exchange pair.

Step 2: Let $((i,j),(u,v))$ be the selected pair of lanes. Compute the payoffs for each carrier:

$$\hat{\pi}^A_{ij,uv} = MC^A_{ij}(L^A) - MC^A_{uv}(L^A \backslash (i,j) \cup (u,v)),$$

$$\hat{\pi}^B_{uv,ij} = MC^B_{uv}(L^B) - MC^B_{ij}(L^B \backslash (u,v) \cup (i,j)),$$

Step 3: If the individual payoffs are nonnegative, i.e., $\hat{\pi}^A \geq 0$ and $\hat{\pi}^B \geq 0$, and the combined payoff is strictly positive, i.e., $\hat{\pi}^A + \hat{\pi}^B > \epsilon$, execute the exchange, update the lane sets of the carriers, and return to Step 1. Otherwise, discard the exchange and stop.

This algorithm terminates after a finite number of iterations because the combined payoff is greater than ϵ at each iteration and the minimum total cost for the carriers is positive. The only computationally intensive step in the algorithm is the ranking of the lanes according to their marginal costs. To compute the marginal cost of a lane, we have to solve two LCPs. Fortunately, LCP can be solved efficiently as a min-cost flow problem.

3.5.2.2 No Information Sharing With Side Payments

In this lane exchange setting, a carrier needs to select a lane to offer and needs to determine an appropriate side payment, or price, for the offered lane. Note that prices for lanes are set before hand to avoid price bargaining. In the lane mechanism we propose, the carriers still offer the lanes with the highest marginal costs (or highest marginal costs per mile). However, as a lane exchange may result in cost savings, the carriers have an incentive to make the offer attractive in order to realize the cost savings. Side payments provide a mechanism for making an offered lane more attractive. We propose a pricing method that is based on the synergies of a lane within the network of the carrier. We can assess the synergies of a lane by comparing the marginal cost to the lane cost. For instance, when the marginal cost of lane (i, j) for Carrier A is equal to $(1 + \theta)c^A_{ij}$, the lane has no synergies within the network of Carrier A, since the marginal cost value implies that Carrier A has to reposition from j to i. On the other hand, when the marginal cost of lane (i, j) for Carrier A is equal to $(1 - \theta)c^A_{ij}$, the lane has perfect synergies within the network of Carrier A, since it implies that covering the lane also serves as an repositioning move for other lanes in the network of Carrier A.

As we increase the price of a lane, i.e., include a side payment, the likelihood of acceptance increases. However, at the same time, the benefits from the lane exchange decrease. As such, lane prices (or side payments) determine how the cost savings that result from exchanging lanes are shared among the carriers. As a carrier does not know these cost savings in advance, since they depend on the other carrier's action, an estimate has to be used. One option is to use $MC_{ij}^A(L^A) - c_{ij}^A$ as an estimate of the cost savings. When the marginal cost of a lane is high compared to the lane cost, the synergies within the network are low. Consequently, the other carrier may be able to cover this lane more cheaply due to potential synergies within his network. Hence, a high price (or side payment) may not be necessary for the other carrier to accept the lane exchange. Sharing the estimated cost savings equally provides a reasonable balance between reducing the risk of a rejection by the other carrier and reducing the size of the benefit of the lane exchange. Thus, the carrier may put forth a price halfway between the marginal cost and original lane cost: $c_{ij}^A + \frac{MC_{ij}^A(L^A) - c_{ij}^A}{2}$.

When the marginal cost of a lane is low compared to the lane cost (but still greater than the marginal cost), there are synergies within the network. Consequently, it is less likely that the other carrier can cover this lane more cheaply as a result of potential synergies within his network. Therefore, the carrier's priority is to make the exchange attractive to the other carrier so as to avoid a rejection of the lane exchange. This means that a higher portion of the estimated cost savings has to be "transferred" to the other carrier to increase the likelihood of acceptance. Thus, the carrier may opt to offer a price that is closer to the marginal cost of the lane (closer than halfway between the lane cost and the marginal cost). We have chosen to use the the adjustment factor $\frac{(1+\theta)c_{ij}^A}{MC_{ij}^A(L^A)}$, which results in the following rule for determining the price:

$$p_{ij}^A = \begin{cases} c_{ij}^A + \frac{(1+\theta)c_{ij}^A}{MC_{ij}^A(L^A)} \times \frac{MC_{ij}^A(L^A) - c_{ij}^A}{2} & \text{if } MC_{ij}^A(L^A) \geq c_{ij}^A, \\ MC_{ij}^A(L^A) & \text{otherwise.} \end{cases}$$

Note that the adjustment factor is a value between 1 and 2 (since $\theta \leq 1$) and that it gets larger when the marginal cost gets smaller.

The following algorithm evaluates the benefits of this exchange mechanism for two carriers A and B, where we assume that the carriers continue to exchange lanes as long as beneficial lane exchanges are identified:

NIWS:

Step 1: Rank the lanes that can be offered according to marginal costs (or per mile marginal costs) for both carriers and select the best one from each to form the lane exchange pair.

Step 2: Let $((i,j),(u,v))$ be the selected pair of lanes. Compute the prices on the lanes:

$$p_{ij}^A = \begin{cases} c_{ij}^A + \frac{(1+\theta)c_{ij}^A}{2\times MC_{ij}^A(L^A)}(MC_{ij}^A(L^A) - c_{ij}^A) & \text{if } MC_{ij}^A(L^A) > c_{ij}^A, \\ MC_{ij}^A(L^A) & \text{otherwise.} \end{cases}$$

$$p_{uv}^B = \begin{cases} c_{uv}^B + \frac{(1+\theta)c_{uv}^B}{2\times MC_{uv}^B(L^B)}(MC_{uv}^B(L^B) - c_{uv}^B) & \text{if } MC_{uv}^B(L^B) > c_{uv}^B, \\ MC_{uv}^B(L^B) & \text{otherwise.} \end{cases}$$

Compute the payoffs for each carrier:

$$\hat{\pi}_{ij,uv}^A = MC_{ij}^A(L^A) - MC_{uv}^A(L^A \backslash (i,j) \cup (u,v)) - p_{ij}^A + p_{uv}^B,$$

$$\hat{\pi}_{uv,ij}^B = MC_{uv}^B(L^B) - MC_{ij}^B(L^B \backslash (u,v) \cup (i,j)) - p_{uv}^B + p_{ij}^A,$$

Step 3: If the individual payoffs are nonnegative, i.e., $\hat{\pi}^A \geq 0$ and $\hat{\pi}^B \geq 0$, and the combined payoff is strictly positive, i.e., $\hat{\pi}^A + \hat{\pi}^B > \epsilon$, execute the exchange, update the lane sets of the carriers, and return to Step 1. Otherwise, discard the exchange and stop.

As before, the algorithm terminates after a finite number of iterations and the only computationally intensive step is computing the marginal cost of the lanes.

3.5.2.3 With Information Sharing - No Side Payments

In this lane exchange setting, the carriers share information to better be able to identify lane exchange opportunities. More specifically, they share their cost structure as well as the lanes they need to serve. With information about the other carrier, a carrier can consider and analyze potential counter offers and estimate the outcome of possible lane exchanges, and choose a strategy accordingly to increase the value of the benefits as well as the probability of acceptance of a lane exchange.

A carrier only needs to select a lane to offer to the other carrier. However, this selection should be based not only on information about their own network, but also on information about other carrier's network and thus expectations about the other carrier's strategy. We propose to model the selection problem as a two-person non-zero sum game where the rows correspond to lanes from Carrier A and columns correspond to lanes from Carrier B. As before, the payoff for a carrier for a given a lane exchange pair is the value of the cost resulting from the exchange. Both carriers can enumerate all possible lane exchange pairs and compute the corresponding payoffs (thus constructing the payoff matrix for the game). Then, based on this information, the carriers can choose the strategy that maximizes the minimum payoff to themselves, which corresponds to finding a Nash equilibrium of the two-person non-zero sum game.

We have the following results:

Lemma 8 *A mixed-strategy Nash equilibrium point always exists for the lane exchange game.*

Proof: Directly follows from the well-known Nash's theorem: "Every finite strategic game has a mixed strategy Nash equilibrium" [30]. □

Lemma 9 *A pure-strategy Nash equilibrium point may not exist for the lane exchange game.*

Proof: Consider the example in Figure 12. Two carriers, namely I and II, have four lanes each, $\{A, B, E, F\}$ and $\{C, D, G, H\}$ respectively and they have identical cost structures on this network (cost figures are shown in Figure 12). Suppose that Carrier

I can offer either lane A or B for an exchange, and Carrier II can offer either lane C or D. Hence, the resulting payoff matrix is as follows:

$$\begin{bmatrix} (2.5, 10.7) & (10.7, 2.5) \\ (10.7, 2.5) & (2.5, 10.7) \end{bmatrix}$$

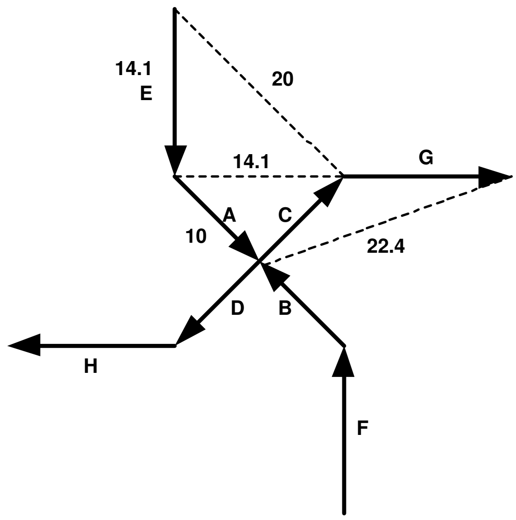

Figure 12: An instance with no pure strategy Nash equilibrium.

Consequently, this instance has no pure strategy Nash equilibrium point. □

Based on these results, we first check whether there exists a pure-strategy Nash equilibrium point for the lane exchange game. If one or more pure-strategy Nash equilibria exist, then we select the one with the highest combined benefit to both carriers. When a pure-strategy Nash equilibrium point does not exist, we find the set of mixed-strategy equilibria and select the one with the highest combined benefit to both carriers. For that mixed-equilibrium point, we select the row and column with the highest probabilities as the strategies for the carriers. Note that, regardless of the type of the equilibrium, this method may not return the best lane exchange pair from a system-wide perspective. This is a result of the selfishness of the carriers. Intuitively, though, this method should be quite effective in terms of the resulting benefits. With information sharing, the carriers can evaluate the response of the other carrier and thus choose a strategy that maximizes the expected cost savings.

The following algorithm evaluates the benefits of this exchange mechanism for two carriers A and B, where we assume that the carriers continue to exchange lanes as long as beneficial lane exchanges are identified:

<u>WINS:</u>

Step 1: For every possible pair of lanes $((i,j),(u,v))$ compute the payoffs for each carrier:

$$\hat{\pi}^A_{ij,uv} = MC^A_{ij}(L^A) - MC^A_{uv}(L^A \backslash (i,j) \cup (u,v)),$$

$$\hat{\pi}^B_{uv,ij} = MC^B_{uv}(L^B) - MC^B_{ij}(L^B \backslash (u,v) \cup (i,j)),$$

$$\pi^A_{ij,uv} = \begin{cases} \hat{\pi}^A_{ij,uv} & \hat{\pi}^A_{ij,uv}, \hat{\pi}^B_{uv,ij} \geq 0,\ \hat{\pi}^A_{ij,uv} + \hat{\pi}^B_{uv,ij} > 0, \\ 0 & \text{otherwise,} \end{cases}$$

$$\pi^B_{uv,ij} = \begin{cases} \hat{\pi}^B_{uv,ij} & \hat{\pi}^A_{ij,uv}, \hat{\pi}^B_{uv,ij} \geq 0,\ \hat{\pi}^A_{ij,uv} + \hat{\pi}^B_{uv,ij} > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Step 2: Construct the payoff matrix using π^A's and π^B's.

Step 3a: Solve the two-person non-zero sum game to find the pure-strategy Nash equilibrium points.

Step 3b: If multiple pure-strategy Nash equilibria exist, then select the one with the highest combined benefits $(\pi^A + \pi^B)$.

Step 3c: If no pure-strategy Nash equilibrium point exists, solve the game to find the mixed-strategy Nash equilibrium points. If multiple mixed Nash equilibria exist, then select the one with the highest combined benefits $(\pi^A + \pi^B)$.

For the selected mixed-strategy Nash equilibrium point, select the row and column strategies with highest probabilities and take corresponding payoffs as the payoffs of the carriers.

Step 4: If the individual payoffs are nonnegative, i.e., $\pi^A \geq 0$ and $\pi^B \geq 0$, and the combined payoff is strictly positive, i.e., $\pi^A + \pi^B > \epsilon$, execute the exchange, update the lane sets of the carriers, and return to Step 1. Otherwise, discard the exchange and stop.

The algorithm terminates after a finite number of iterations. This algorithm is much more computationally intensive. To construct the payoff matrix, we have to compute the payoffs for all possible pairs lane exchange pairs. For a given lane exchange pair, this involves solving four LCPs (two for each marginal cost calculation). After constructing the payoff matrix, we have to find the Nash equilibria of the game.

3.5.2.4 With Information and With Side Payments

As mentioned before, by allowing side payments the number of lane exchange opportunities increases. Observe that with information sharing, but without side payoffs, if for a given lane exchange pair one of the payoffs is negative, the exchange will not be acceptable. Suppose that the reason for one of the payoffs being negative is that the offered lane is a high cost lane for the receiving carrier, even though there are more synergies within the network of the receiving carrier than there are synergies within the network of the offering carrier. On the other hand, the payoff for the offering carrier may be quite high due to the fact that he is giving away a high cost lane. A side payment may result in nonnegative payoffs for both carriers.

In this lane exchange setting, a carrier needs to select a lane to offer and needs to determine an appropriate side payment, or price, for the offered lane. Again, we propose to model the decision process as a two-person non-zero sum game where the rows correspond to lanes from Carrier A and columns correspond to lanes from Carrier

B. The only difference is that the payoff values have to include any side payments. Determining appropriate side payments, however, is quite complicated. The basic idea of a side payment is to make an offer attractive to the other carrier. However, when a carrier includes a side payment with a lane, it affects the payoffs of all the lane exchange pairs that involve the lane. Furthermore, the other carrier may include a side payment with the offered lane as well, in effect returning a portion of the side payment. We propose to use a "conservative approach" for determining prices. If Carrier A offers lane (i,j) and Carrier B would choose (u,v) as the counter-offer to maximize his payoff, i.e., $(u,v) = argmax_{(u,v)\in L^B}\hat{\pi}^B_{uv,ij}$, but $\hat{\pi}^B_{uv,ij} < 0$, then Carrier A will price (i,j) at $p^A_{ij} = \epsilon - \hat{\pi}^B_{uv,ij}$ resulting in a positive payoff for Carrier B. Of course, Carrier A will do so only if $\hat{\pi}^A_{ij,uv} + \hat{\pi}^B_{uv,ij} > \epsilon$ otherwise his own payoff would become negative.

The following algorithm evaluates the benefits of this exchange mechanism for two carriers A and B, where we assume that the carriers continue to exchange lanes as long as beneficial lane exchanges are identified:

WIWS:

Step 1: For every possible lane exchange pair $((i,j),(u,v))$ compute the tentative payoffs for each carrier:

$$\bar{\pi}^A_{ij,uv} = MC^A_{ij}(L^A) - MC^A_{uv}(L^A\backslash(i,j)\cup(u,v)),$$

$$\bar{\pi}^B_{uv,ij} = MC^B_{uv}(L^B) - MC^B_{ij}(L^B\backslash(u,v)\cup(i,j)),$$

Step 2: For every lane (i,j) of Carrier A, determine $(u,v) = argmax_{(u,v)\in L^B}\hat{\pi}^B_{uv,ij}$ and

$$p^A_{ij} = \begin{cases} \epsilon - \bar{\pi}^B_{uv,ij} & \bar{\pi}^B_{uv,ij} < 0,\ \bar{\pi}^A_{ij,uv} + \bar{\pi}^B_{uv,ij} > \epsilon, \\ 0 & \text{otherwise.} \end{cases}$$

For every lane (u, v) of Carrier B, determine $(i, j) = argmax_{(i,j)\in L^A}\hat{\pi}^A_{ij,uv}$ and

$$p^B_{uv} = \begin{cases} \epsilon - \bar{\pi}^A_{ij,uv} & \bar{\pi}^A_{ij,uv} < 0,\ \bar{\pi}^A_{ij,uv} + \bar{\pi}^B_{uv,ij} > \epsilon, \\ 0 & \text{otherwise.} \end{cases}$$

Update the tentative payoff values as

$$\hat{\pi}^A_{ij,uv} = \bar{\pi}^A_{ij,uv} - p^A_{ij} + p^B_{uv},$$

$$\hat{\pi}^B_{uv,ij} = \bar{\pi}^B_{uv,ij} + p^A_{ij} - p^B_{uv}.$$

For every possible lane exchange pair $((i, j), (u, v))$ compute the resulting payoffs

$$\pi^A_{ij,uv} = \begin{cases} \hat{\pi}^A_{ij,uv} & \hat{\pi}^A_{ij,uv}, \hat{\pi}^B_{uv,ij} \geq 0,\ \hat{\pi}^A_{ij,uv} + \hat{\pi}^B_{uv,ij} > 0, \\ 0 & \text{otherwise,} \end{cases}$$

$$\pi^B_{uv,ij} = \begin{cases} \hat{\pi}^B_{uv,ij} & \hat{\pi}^A_{ij,uv}, \hat{\pi}^B_{uv,ij} \geq 0,\ \hat{\pi}^A_{ij,uv} + \hat{\pi}^B_{uv,ij} > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Step 3: Construct the payoff matrix using π^A's and π^B's.

Step 4a: Solve the two-person non-zero sum game to find the pure-strategy Nash equilibrium points.

Step 4b: If multiple pure-strategy Nash equilibria exist, then select the one with the highest combined benefits $(\pi^A + \pi^B)$.

Step 4c: If no pure-strategy Nash equilibria exists, solve the game to find the mixed-strategy Nash equilibrium points. If multiple mixed-strategy Nash equilibria exist, then select the one with the highest benefits $(\pi^A + \pi^B)$. For the select mixed-strategy Nash equilibria, select the row and column strategies with highest probabilities and take corresponding payoffs as the payoffs of the carriers.

Step 5: If the individual payoffs are nonnegative, i.e., $\pi^A \geq 0$ and $\pi^B \geq 0$, and the combined payoff is strictly positive, i.e., $\pi^A + \pi^B > \epsilon$, execute the exchange, update the lane sets of the carriers, and return to Step 1. Otherwise, discard the exchange and stop.

This algorithm terminates after a finite number of iterations and has the same computational challenges as the previous one.

All proposed algorithms require the solution of a large number of LCPs to compute the marginal costs of the lanes of the carriers. The following property concerning the marginal cost of a lane allows us to significantly reduce the computational time of our algorithms.

Theorem 4 *The marginal cost of a lane is less than or equal to the value of the corresponding dual variable in an optimal solution to the dual of the lane covering problem.*

Proof: Let L^k be the lane set of carrier k and r_{ij}^k be the number of times lane $(i,j) \in L^k$ is required to be covered. Hence, for a carrier k, the dual of the lane covering problem is as follows:

$$\begin{aligned}
DLCP^k \quad z(L^k) &= \max \sum_{(i,j)\in L^k} r_{ij}^k a_{ij}^k && (45)\\
s.t.\ a_{ij}^k + y_i^k - y_j^k &= c_{ij}^k \ \ \forall (i,j) \in L^k && (46)\\
y_i^k - y_j^k &\leq \theta c_{ij}^k \ \ \forall (i,j) \in A && (47)\\
a_{ij}^k &\geq 0 \ \ \forall (i,j) \in L^k. && (48)
\end{aligned}$$

Let $(\hat{a}^k, \hat{y}^k)$ represent the optimal solution to $DLCP^k$. As the optimal objective function value of the dual LCP is equal to the total cost of covering all the lanes by Carrier k, the marginal cost of lane (u,v) to Carrier k's network is equal to:

$$MC_{uv}^k(L^k) = z^*(L^k \cup (u,v)) - z^*(L^k)$$

where $z^*(.)$ represents the optimal objective function value of the associated dual LP. After adding lane (u, v) to the network, the resulting cost of Carrier k can be determined by solving the modified dual problem:

$$
\begin{aligned}
\overline{DLCP}^k \quad z(L^k \cup (u,v)) &= \max \sum_{(i,j)\in L^k} r_{ij}^k a_{ij}^k + a_{uv}^k && (49)\\
s.t.\ a_{ij}^k + y_i^k - y_j^k &= c_{ij}^k \ \ \forall (i,j) \in L^k && (50)\\
a_{uv}^k + y_u^k - y_v^k &= c_{uv}^k && (51)\\
y_i^k - y_j^k &\leq \theta c_{ij}^k \ \ \forall (i,j) \in A && (52)\\
a_{ij}^k &\geq 0 \ \ \forall (i,j) \in L^k \cup (u,v). && (53)
\end{aligned}
$$

Let $(\bar{a}^k, \bar{y}^k)$ represent the optimal solution to $\overline{DLCP}^k$. As $(\bar{a}^k, \bar{y}^k)$ is feasible for the original dual problem $DLCP^k$, we have

$$\sum_{(i,j)\in L^k} r_{ij}^k \bar{a}_{ij} \leq \sum_{(i,j)\in L^k} r_{ij}^k \hat{a}_{ij}$$

since $DLCP^k$ is a maximization problem. Combining these two results, we obtain the following property for the marginal costs:

$$MC_{uv}^k(L^k) = z^*(L^k \cup (u,v)) - z^*(L^k) = \sum_{(i,j)\in L^k} r_{ij}^k \bar{a}_{ij} + \bar{a}_{ij}^k - \sum_{(i,j)\in L^k} r_{ij}^k \hat{a}_{ij} \leq \bar{a}_{ij}^k . \square$$

The optimal values of the dual variables can easily be found using linear programming technology. Consequently, we have an efficient mechanism for obtaining upper bounds on the marginal costs of the lanes, which we will use instead of the marginal costs unless these values are insufficient. For example, the lane exchange mechanism without information sharing and without side payments requires the identification of the lane with the highest marginal cost. We sort the lanes based on the upper bounds on the marginal cost and compute the actual marginal cost for the lane with the largest upper bound, which provides a lower bound on the value of the highest marginal cost. For the remaining lanes, if the upper bound is less than or equal to the lower bound on the value of the highest marginal cost, we simply skip the computation of the true marginal cost. Otherwise, we compute the true marginal cost of the lane and update the lower bound if necessary. Computational experience shows that this simple procedure significantly reduces computation times. The lane exchange mechanism with information sharing but without side payments can take advantage of Theorem 4. The computation of the tentative payoffs ($\hat{\pi}$'s) involves two marginal costs. One of these can be replaced with its upper bound to get an upper bound on the payoff. Only if the upper bound on the payoff is positive, do we compute the actual marginal cost of the lane for which we have used the upper bound. Let $\bar{a}_{ij}^A$ and $\bar{a}_{uv}^B$ represent the relevant upper bounds, then we examine

$$\tilde{\pi}_{ij,uv}^A = \bar{a}_{ij}^A - MC_{uv}^A(L^A \cup (u,v))$$

and

$$\tilde{\pi}_{uv,ij}^B = \bar{a}_{uv}^B - MC_{ij}^B(L^B \cup (i,j)).$$

to see if true payoff values need to be computed. The same idea can be applied in case there is information sharing and there are side payments. Computational experience shows significant reductions in computation times.

3.5.3 Computational Study

We have carried out a computational study to evaluate the performance of the proposed lane exchange mechanisms.

Random instances are generated as before. All experiments assume there are two carriers with identical cost structures, which is the setting with the least potential for cost savings opportunities. To investigate the impact of lanes that have to be served by the carrier, we create 4 new instances for every instance by adding lanes that have to be served by a carrier. (Thus, the number of lanes that can be exchanged is the same for the 4 instances.) The 4 instances are created by adding 0%, 50%, 100%, and

200% of the number of lanes that can be exchanged. That means that if the original instance has 120 lanes, the variant with 50% of lanes that have to be served by a carrier has a total of 180 lanes, i.e., 120 that can be exchanged (assigned randomly to the two carriers) and 60 that cannot be exchanged (assigned randomly and evenly to the two carriers).

Our goal is to show that bilateral lane exchange mechanisms are capable of identifying and exploiting the synergies that exist among the lanes of the two carriers. We determine the maximum possible savings by computing the minimum cost for Carrier A (obtained by solving LCP over the lanes assigned to Carrier A) plus the minimum cost for Carrier B (obtained by solving LCP over the lanes assigned to Carrier B) minus the cost of a centralized solution (obtained by solving MCLCP over all lanes). We use the term optimality gap to refer to the maximum possible savings as a percentage of the cost of the perfect collaboration.

We evaluate each of the four lane exchange mechanisms (NINS, NIWS, WINS, and WIWS) by computing the optimality gap after a lane exchange mechanism has been applied. We solve the two-person non-zero sum games in mechanisms WINS and WIWS using Gambit software package [34]. Table 6 summarizes the performance of the exchange methods. We report the optimality gap for the random assignment (Random), the average optimality gap for mechanism NINS (NINS), the average computation time for mechanism NINS in CPU seconds (cpu NINS), the average optimality gap for mechanism NIWS (NIWS), the average computation time for mechanism NIWS (cpu NIWS), the average optimality gap for mechanism WINS (WINS), the average computation time for mechanism WINS (cpu WINS), the average optimality gap for mechanism WIWS (WIWS), and the average computation time for mechanism WIWS (cpu WIWS). For each instance size, we present average statistics for the different levels of lanes that have to be served by the carrier itself.

Table 6: Optimality gaps for the proposed lane exchange mechanisms.

40 lanes	**Random**	**NINS**	**cpu NINS**	**NIWS**	**cpu NIWS**	**WINS**	**cpu WINS**	**WIWS**	**cpu WIWS**
0%	5.66	3.55	3.20	3.14	2.00	0.37	74.60	0.16	268.00
50%	8.07	4.76	3.00	3.81	2.40	0.92	69.40	0.36	302.20
100%	5.45	4.50	2.80	4.69	1.80	0.77	56.80	0.26	239.60
200%	11.12	8.03	2.60	7.48	1.80	1.97	82.00	0.73	352.20
80 lanes	**Random**	**NINS**	**cpu NINS**	**NIWS**	**cpu NIWS**	**WINS**	**cpu WINS**	**WIWS**	**cpu WIWS**
0%	3.34	2.28	3.80	2.29	2.40	0.11	233.20	0.07	1034.00
50%	5.31	4.52	3.20	4.07	1.60	0.34	277.40	0.20	1345.40
100%	5.48	4.79	3.00	4.71	1.40	0.70	270.60	0.55	1233.00
200%	4.42	4.42	2.20	4.21	1.20	0.69	224.40	0.31	1192.60
120 lanes	**Random**	**NINS**	**cpu NINS**	**NIWS**	**cpu NIWS**	**WINS**	**cpu WINS**	**WIWS**	**cpu WIWS**
0%	3.27	2.61	3.20	2.61	1.40	0.13	697.40	0.03	3042.60
50%	3.86	2.03	3.80	1.85	2.60	0.12	596.60	0.13	2747.40
100%	4.96	3.59	3.40	3.20	2.40	0.13	879.00	0.09	3714.80
200%	5.61	4.31	3.20	4.54	1.60	0.40	817.20	0.29	4360.00
Average	5.54	4.12	3.12	3.88	1.88	0.55	356.55	0.27	1652.65

The computational experiments reveal that the proposed lane exchange mechanisms perform quite well and are capable of identifying and exploiting the synergies among the lanes of the two carriers. The average optimality gap for $NINS$ is 4.12% and the average optimality gap for of $NIWS$ is 3.88%, which correspond to a 26% and 30% reduction of the optimality gap for the random assignment of the lanes to the carriers. Although $NIWS$ outperforms $NINS$ on average, for some instances $NINS$ has lower optimality gap than $NIWS$. As we mentioned earlier, allowing side payments is not always beneficial and here we observe this phenomenon in our experiments.

It is also clear that the lane exchange mechanism with information sharing perform considerably better than those without information sharing. The average optimality gap for $WINS$ is 0.55% and the average optimality gap for $WIWS$ is only 0.27%, which correspond to a 90% and a 95% reduction from the optimality gap of random assignment of the lanes to the carriers. In fact, in most of the instances $WIWS$ achieves the optimal solution (i.e., the centralized solution representing perfect collaboration). Information sharing allows carriers to select their strategies based on the anticipated counter offers and thus allowing carriers to identify the best possible lane exchange. Allowing side payments helps too, but its effect is much less significant than sharing information.

Next, we generate 60 instances in which the sets of lanes that have to served by a carrier are geographically separated. This is accomplished by dividing the square region in an upper triangular area and the complementing lower triangular area and generating the lanes that have to be served by one carrier in the upper triangular region and the lanes that have to be served by the other carrier in the lower triangular region. The results can be found in Table 7.

Table 7: Optimality gaps for the proposed lane exchange mechanisms when the carriers have geographically separated sets of lanes that they have to serve.

40 lanes	**Random**	**NINS**	**cpu NINS**	**NIWS**	**cpu NIWS**	**WINS**	**cpu WINS**	**WIWS**	**cpu WIWS**
50%	5.23	4.10	3.20	4.10	1.20	0.70	64.00	0.34	277.40
100%	6.64	4.63	2.80	4.37	2.00	1.32	60.20	0.54	229.40
200%	10.15	6.91	3.00	5.70	2.00	2.29	63.40	0.59	236.60
80 lanes	**Random**	**NINS**	**cpu NINS**	**NIWS**	**cpu NIWS**	**WINS**	**cpu WINS**	**WIWS**	**cpu WIWS**
50%	10.55	4.47	4.60	4.12	3.00	0.24	543.40	0.27	1651.00
100%	11.27	5.04	4.40	3.93	4.40	0.41	503.00	0.32	1549.80
200%	7.37	5.25	3.40	4.28	2.80	0.63	368.60	0.43	1606.00
120 lanes	**Random**	**NINS**	**cpu NINS**	**NIWS**	**cpu NIWS**	**WINS**	**cpu WINS**	**WIWS**	**cpu WIWS**
50%	5.81	2.87	5.40	2.56	4.00	0.19	1094.20	0.10	3677.80
100%	4.19	3.21	4.00	2.98	2.40	0.29	720.60	0.07	3864.20
200%	4.67	3.39	3.40	3.26	2.00	0.40	690.60	0.24	3987.00
Average	7.32	4.43	3.80	3.92	2.64	0.72	456.44	0.32	1897.69

For these instances, the average optimality gap for $NINS$ and $NIWS$ is 4.43% and 3.92% respectively. These results correspond to a 40% and a 47% reduction of the optimality gap of random assignment of the lanes to the carriers, hence the performance is better than when the lanes that have to be served by the carriers are randomly assigned to carriers and have no special geographical properties. The average optimality gap for $WINS$ and $WIWS$ are 0.72% and 0.32%, respectively, and these results correspond to a 90% and 96% reduction from the optimality gap of the random assignment of the lanes to the carriers. Comparing these results to the previous results, it appears that the performance of the lane exchange process is better when the carriers have a "clear identity" in the form of geographically separated and clustered sets of lanes that have to be served by them. This is probably due to the fact that fewer synergies arise with such sets.

The computational time of the exchange mechanism are within the acceptable limits. Although the average computation time of the proposed mechanisms increases as the mechanism becomes more complex and so effective, the highest computation time is less than 4500 seconds, which indicates that the proposed mechanisms can be implemented in practice.

CHAPTER IV

ALLOCATING COSTS IN INVENTORY ROUTING PROBLEM

4.1 Introduction

Vendor managed inventory is a coordinated effort between a supplier and its customers to optimize the performance of a supply chain. In VMI, the supplier takes the responsibility of managing inventories at the customer locations. In a conventional inventory management setting, each customer monitors its own inventory level, places orders to the supplier when required and controls the timing and size of the orders being placed. In VMI, the supplier monitors the inventory levels at the customer locations, generates the order for a customer when required and makes the decisions related to when to deliver, how much to deliver, and how to deliver to its customers.

VMI offers various benefits to the supplier and its customers. Under a VMI setting, the supplier has access to true point of sale data as well as the exact inventory levels at the customer locations. With this information, the supplier can better schedule its production and distribution and plan deliveries more accurately to minimize the stockouts at customer locations. Furthermore, reduced distortion of demand data results in decreased inventory levels along the supply chain and reduced inventory holding costs for both the supplier and its customers. Also, because of the increased flexibility in scheduling the time and volume of the deliveries, the supplier can decrease its transportation costs with better coordination of deliveries.

To realize the benefits of VMI due to better coordination, the supplier and its customers have to overcome various challenges. One of these challenges faced by the supplier is in developing a distribution plan that minimizes the total distribution and inventory holding costs without causing product shortage at customer locations. In this problem, referred as the *inventory routing problem (IRP)* in the literature, the challenge is to coordinate the routing decisions with the inventory management decisions. That is, the supplier has the opportunity to exploit synergies between customers to reduce distribution costs by serving nearby customers on the same route at the same time while making sure that the inventory holding and product shortage costs also stay reasonable. In order to do this, the supplier has to make simultaneous decisions on vehicle routes, route frequencies, and volume delivered to each customer using these routes.

Solving IRP's are challenging due to the interacting effects of the depot and customer locations, fleet size, vehicle and warehouse capacities, customer consumption rates, time restrictions on the deliveries and planning horizon length on the supplier's routing/volume decisions. For example, in contrast to the vehicle routing problem, in IRP decisions are made over a planning horizon and decisions made in one period have impacts on decisions in consecutive periods. As the supplier makes the decisions on when and on which route(s) to serve customers and how much to deliver to customers the number of feasible distribution strategies becomes practically endless.

Our research is motivated by our relationship with a large industrial gas company that operates a vendor managed inventory resupply policy. The company replenishes the storage tanks at customer locations by a homogenous fleet of tanker trucks under the company's control. The customers and trucks are assigned to a specific facility, and tanker trucks and storage tanks contain only a particular type of product. The company incurs the inventory holding costs at the production facility and the customer sites, so we can ignore inventory holding costs. Therefore, the inventory management component affects the feasibility of the distribution plan rather than its cost. We also assume that the consumption rates of customers are deterministic and stationary over time. As a result, the problem is reduced to minimizing the total cost of transportation of a single product from a single depot to a set of customers with deterministic and stationary demand over a planning horizon (possibly infinite).

In a practical setting another challenge faced by the supplier is to calculate the "cost-to-serve" for each customer. While this information is of value to the supplier for many purposes including marketing and sales efforts, it is hard to identify a simple formula that allocates the total distribution costs among its customers. Simple allocation methods that distribute the costs proportional to customer distance to the depot, storage capacity and consumption rate but that do not take into account the synergies between the customers do not represent the true cost-to-serve and may result in over or under charging the customers. In this chapter we address this challenge and develop methods for the supplier to calculate a cost-to-serve for each customer or allocate the total distribution costs among the customers. There already exists a large body of literature proposing solution approaches to variations of the IRP and we take these as given.

True cost-to-serve information is of value to the supplier for marketinng/sales purposes and has effects in distribution planning. First of all, this information may be used in contracting and pricing stages. If the price charged to a customer for distribution does not represent the actual cost of serving that customer the supplier may lose existing customers to competitors or decline profitable prospective customers. The supplier may also use this information to revise the previously given routing decisions by comparing the actual cost of a customer to the price charged to that customer during the contracting stage. That is, if an actual costs of a set of customers is considerably greater than the total price charged to them, it may be an indication of a possible improvement to the initial IRP solution. Also, even if the company prefers to use the original pricing scheme to charge the customers, the actual cost information may be employed to identify the low cost/high profit customers to prioritize in case of a product shortage. Finally, the marketing department may use the information to identify regions to seek potential customers from to increase the overall profitability of an IRP route.

In general determining the cost-to-serve is not trivial in distribution problems. If each customer is served individually, then allocating distribution costs to the customers is easy. Simple cost allocation methods (including those that are used in current practice) that ignore the synergies that may exist between customers assume that each customer is served individually and allocate costs based on individual attributes of the customers such as distance to the depot, storage capacity and consumption rate. When multiple customers are visited together on delivery routes, the situation becomes much more complex. That is, it is not simple to determine the proportion of the delivery route cost each customer is responsible for. Although geographical locations, storage capacities and consumption rates still affect the cost allocations of the customers, the interaction of these parameters is the main factor in determining the cost-to-serve values.

We illustrate these concepts over a simple example with three customers A, B, C and a depot D as shown in Figure 13(a). We assume that all the customers have a storage tank size equal to half the size of the capacity of the delivery truck. Their consumption rates are equal and the truck is able to serve the customers without any shortage over the planning horizon. Finally the customers have the same distance to the depot. A pricing scheme that ignores the synergies will assign the same cost to each customer because they are identical in terms of distance to the depot, storage capacity and consumption rate. However, by bundling customers A and B together, the supplier can serve both customers on the same route. Therefore, the supplier can serve customer A or B for a considerably less cost compared to customer C due to the geographical synergy between A and B.

If the supplier charges these customers according to the pricing scheme discussed above, it risks to lose customer A and B to a competitor, which may serve both of

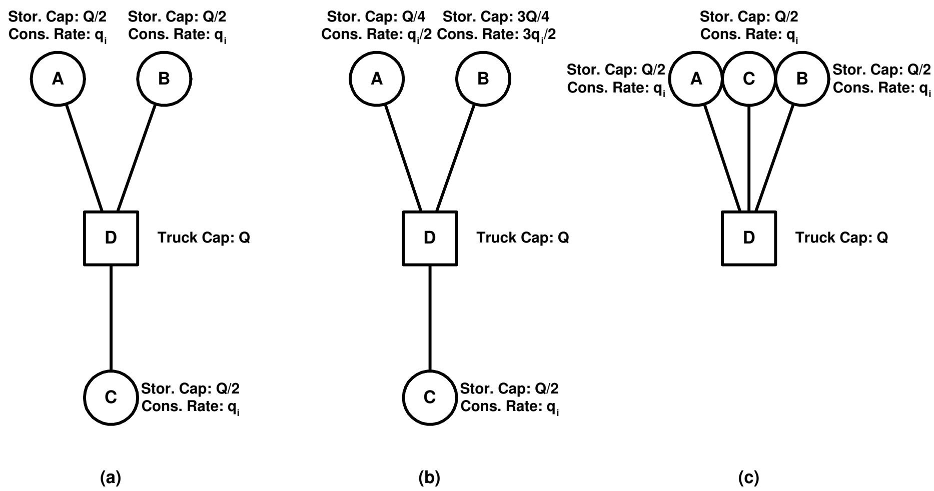

Figure 13: A basic instance.

them at a lower price. Also, in case of a product shortage, the supplier may prioritize the profitable customers A and B instead of customer C. Finally, the supplier may realize that it can increase the profitability of customer C by pairing this customer with a new customer from the same region.

Consider modifying the example slightly and by decreasing the storage capacity and consumption rate of customer A to half of the previous values, and increasing the storage capacity and consumption rate of customer B to one and a half the previous values (shown in Figure 13(b)). As a result, the storage capacity of customer A is a quarter of the size of the capacity of the truck and the storage capacity of customer B is three quarters of the size of the capacity of the truck. The supplier can still serve both customers on the same route by delivering a quarter of the truck capacity to A and remaining to B. However, it is not clear that how the cost of this route will be allocated among customer A and B. Allocating the cost of the route to customers A and B proportional to the delivery volume, hence proportional to 1 and 3 respectively, seems to be a reasonable solution. On the other hand, the fact that they would be responsible for the whole distribution cost if they were served alone suggests that the cost allocations should be equal.

Now consider modifying the original example slightly and changing the location of customer C as in Figure 13(c). Suppose that at most two customers can be served together by a single truck because of the consumption rates and planning period interactions. In this case, any two customers can be served together on a delivery route in a cost efficient way. However, grouping two customer together and allocating the cost of the joint delivery route to these customers and allocating twice this amount to the third customer may not be a rational cost allocation, as all three customers are identical in every single aspect.

In these examples, we illustrate that it is not trivial to allocate the distribution costs among the customers. We also demonstrate that not only the synergies among the customers along a delivery route (intra-route synergies) but also the synergies among the customers on different delivery routes (inter-route synergies) are important in determining the cost-to-serve values of the customers.

The rest of this chapter is organized as follows. In Section 4.2, we review the related work in the literature and briefly discuss cost allocation properties of well studied allocation methods from cooperative game theory. In Section 4.3, we first give a formal statement of the IRP and then discuss solution methods that provides upper and lower bounds as well as an approximation to the optimal objective function of the IRP that will be used in calculating cost-to-serve methods. In Section 4.4, we introduce the inventory routing game (IRG) and develop several cost allocation methods. We also discuss the pros and cons of each cost allocation method that we propose and demonstrate the performance of each method on the constructed examples in this section. In Section 4.5, we computationally demonstrate how our methods perform on randomly generated instances. Concluding remarks are provided in Section 4.6.

4.2 Literature Review

In this section, we review the related work in the literature, which can be categorized into two streams: literature on the inventory routing problem and cooperative game theory.

As stated before, there exists a large body of literature on inventory routing problems. We refer the reader to surveys Campbell et al. [8], Cordeau et al. [10] and Bertazzi et al. [5] for solution approaches on various inventory routing problems. Next we discuss the work by Song and Savelsbergh [40] in detail since it will be referred later in the chapter. In this work, they develop a methodology based on first generating feasible delivery routes or patterns and then use a linear programming problem to select good patterns to obtain a lower bound on the total transportation cost needed to satisfy customer demand over the planning horizon. They prove that it is sufficient to generate only a subset of the patterns to find the optimal solution to the pattern selection linear program and this reduces the computational time considerably.

Another stream of literature that is relevant to our work is on cooperative game theory. Cooperative game theory studies the class of games in which selfish players form collaborations to obtain greater benefits. The cost allocation problem of the VMI we consider in this chapter is a cooperative game. In a cooperative game a common cost (or equivalently benefit) must be allocated among the participating players. The set $N = \{1, \ldots, n\}$ represents the set of players and every $S \subset N$ is called a *coalition* of players, whereas N is called the *grand coalition.* Every cooperative game has an associated characteristic function, $c(.)$ and the cost incurred by a coalition S is represented by $c(S)$. The total cost to be allocated among the players is represented by $c(N)$. Even though, the customers do not form coalitions in our VMI setting, developing a fair mechanism to compute cost-to-serve for each customer corresponds to developing a cost allocation method for the *inventory routing game* (IRG). In this inventory routing game (a formal definition will be presented in Section 4.4), the customers represent the players and the optimal objective function value of the IRP is the characteristic function value of the game.

Next, we briefly summarize the concepts we will use from cooperative game theory literature. In a *budget balanced* cost allocation, the total cost allocated to the participating players is equal to the total cost incurred by the grand coalition, $c(N)$. In a *stable* cost allocation, the total cost allocated to a subset of players should be less than or equal to the total cost incurred by that subset, $c(S)$. The set of cost allocations that are budget balanced and stable is called the *core* of a collaborative game. For a collaborative game, the *core* may be empty, that is, a budget balanced and stable cost allocation may not exist.

Nucleolus, introduced by [37], is the cost allocation that lexicographically maximizes the minimal gain, the difference between the stand alone cost of a subset and its allocated cost, over all the subsets of the collaboration. If the core is non-empty, the nucleolus is included in the core. Nucleolus may still exist when the core is empty if there exists a cost allocation that is budget balanced and the cost it allocates to each player is less than or equal to their stand alone costs. Another well-known cost allocation method is the *Shapley Value*, which is defined for each player as the weighted average of the player's marginal contribution to each subset of the collaboration [38]. Even if the core is non-empty, a cost allocation based on the Shapley Value may not be included in the core.

As general references on cooperative game theory, we refer to Young [48] for a thorough review of basic cost allocation methods and Borm et al. [7] for a survey of cooperative games associated with operations research problems.

Cost allocation in routing and distribution problems has rarely been addressed in the literature. To the best of our knowledge, there is no previously published work on allocating cost of a inventory routing problem. However, there exist closely related studies on the Traveling Salesman (TSG) and Vehicle Routing (VRG) games. The

TSG considers the problem of allocating the cost of a round trip among the cities visited. Depending on the game, there may exist a home city which is not included in the set of players. Tamir [43] considers a TSG with a home city and proves that under symmetric cost structure TSG always has a non-empty core when the number of players are less than or equal to four. Kuipers [26] extends this result to TSG with five players. Note that the core may be empty when the number of players are greater than or equal to six.

Potters et al. [33] discuss TSG and its variant with fixed routes. For the fixed route TSG where for all coalitions the cities are visited in the same order, they prove that the game has a non-empty core if the fixed route solution is the optimal solution to the original TSP and the cost matrix satisfies the triangle inequality. They also introduce a class of matrices that generates TSG with non-empty cores.

Engevall et al. [15] consider the cost allocation problem of a distribution system of a gas and oil company where the total cost of a tour is to be allocated among the customers that are visited. Besides the well-known concepts from cooperative game theory, such as the nucleolus and the Shapley value, they introduce "demand nucleolus", which is similar to the nucleolus except that the excess of a coalition is multiplied by the demand of that coalition. By doing this, they aim to reduce the importance of coalitions that have relatively larger demands in computing the nucleolus.

Faigle et al. [19] also consider the problem of allocating the cost of an optimal traveling salesman tour in a fair way with a cost matrix satisfying triangular inequality and present examples with empty core even with Euclidean distances. By using the optimal value of the Held-Karp relaxation of the TSP, they claim to obtain cost allocations from moat packings in polynomial time using LP duality, guaranteeing that no coalition pays more than $\frac{4}{3}$ times its stand alone cost.

The VRG considers the problem of allocating the cost of optimal vehicle routes among the customers served. Gothe-Lundgren et al. [21] consider the VRG with a homogenous fleet and present conditions when the core of a VRG is non-empty. Then, they use a constraint generation approach to compute the nucleolus of a VRG with a non-empty core. Engevall et al. [16] extend this result to a VRG with heterogeneous vehicles and with some simplifications they determine whether the core is empty or not in a heterogeneous vehicles setting. They finally use the constraint generation approach to compute the nucleolus when the core is non-empty.

There is a close relationship between the core and linear programming duality and a substantial amount of literature exists on using the dual of a problem to find or approximate the cost allocations in the core. Pal and Tardos [32] convert a primal-dual algorithm into a group strategyproof cost-sharing mechanism and develop approximate budget balanced cost allocation methods for two NP-complete problems: metric facility location and single source rent-or-buy network design. Pal and Tardos [32] also mention that for covering games, the core is non-empty if and only if the linear relaxation of the game-defining IP has no integrality gap. Goemans and Skutella [20] establish a similar relationship between the existence of the core and LP duality for several variants of the facility location problem. They show that a core cost allocation exists if and only if the integrality gap is zero for a corresponding linear programming relaxation.

Cost allocation methods studied in the literature for related problems, such as TSG and VRG, do not have to deal with the time or inventory components of the problem and hence cannot be applied directly to IRG. Furthermore, the generic methods proposed for the associated routing games, such as the nucleolus and the Shapley Value, have exponential computation times. In this study, we develop cost allocation methods that have low computational times and are effective in terms of game theoretical and practical aspects.

4.3 Inventory Routing Problem

In this section, we first define the *inventory routing problem* (IRP) that seeks to identify the best distribution strategy to serve customers over a planning horizon and list our assumptions. Next, we discuss existing and new methods that provide feasible solutions and upper and lower bounds on the optimal objective function value of the IRP.

4.3.1 Problem Definition

In this chapter, we consider a set of customers $N = \{1, \dots, n\}$ that receive a single product from a single facility (depot). The IRP is defined on graph G, with vertex set $V = \{0, 1, \dots, n\}$ where 0 represents the depot, and edge set $E = \{(i, j) : i, j \in V\}$. The cost of traveling along an edge (i, j) is denoted by c_{ij} and these cost figures are assumed to satisfy the triangular inequality. The supplier has an infinite supply and is responsible for delivering to customers over a given planning horizon of length T. We assume that all the events occur at discrete time intervals.

Customer i has a finite storage capacity C_i and consumes the product at a deterministic and stationary consumption rate of q_i per period. The initial inventory of customer i is denoted by I_{i0} and the inventory at the end of period t is denoted by I_{it}. The supplier serves customers by executing a set of delivery routes that start and end at the depot. Each route is assigned one vehicle from an infinite number of homogenous vehicles with capacity Q. A feasible delivery route corresponds to an ordered list of customers and the amount of product delivered to each customer on the list such that the total volume delivered does not exceed the truck capacity. The cost of a delivery route is the total transportation cost of the vehicle. Then the IRP seeks to identify a set of minimum cost delivery routes over a given planning horizon such that no stock-outs occur and the storage capacity at each customer location is not exceeded. In other words the supplier has to choose his distribution strategy over a planning period by specifying which routes to execute and how much to deliver to each customer on a given route and when these routes depart from the facility.

Next, we summarize the rest of our assumptions. First, we assume that all the routes to be executed during a period are performed at the beginning of the period and completed instantly. The demand at each customer is realized after the deliveries are made at the beginning of each period. The cost of executing a route only depends on the travel costs on the arcs, c_{ij}'s, and does not depend on the amount of product transported or delivered. No shortages at the customer locations are allowed. Deliveries are not restricted by customer specified time windows. Storage capacities at customer locations affect the feasibility of a distribution strategy, however, since we assume that the supplier incurs all the inventory holding costs, we ignore these costs in formulating the IRP.

4.3.2 Solution Methodologies for the IRP

In this subsection, we discuss several methods for finding or approximating optimal solutions for the IRP.

As discussed in the literature, solving the inventory routing problem to optimality is a complex task, even for very small instances of the problem. Many techniques have been proposed depending on the various problem settings. Although these approaches may take into account many complexities of the IRP and yield high quality solutions, they are quite complicated and time consuming to use as a characteristic function of a game. Our focus in this work is to compute cost-to-serve values for the customers rather than solving the IRP, hence we employ quick mechanisms with possibly simplifying assumptions to approximate the optimal solution value instead of using a method with high level of detail and long solution time. Therefore, we propose several methods with simplifying assumptions to obtain approximations and lower and upper bounds to the optimal IRP solution. We propose four different methods, two of them serve as lower bounds, one of them serves as an upper bound and the last one is an approximation to the optimal objective function of the IRP.

4.3.2.1 Mixed Integer Programming Model

First, we construct a mixed integer programming model to solve the IRP we consider in this chapter. The solution to this model provides the optimal distribution strategy for the IRP faced by the supplier.

Before introducing the model, we first define the terminology that will be used in the formulation. A "*delivery pattern*" represents a route with a set of customers to be visited in order. The cost of a pattern p is denoted by θ_p and is equal to the optimal cost of the traveling salesman problem that visits the depot and the customers included in the given set. Let $\mathcal{P}$ be the set of all delivery patterns and let a_{ip} be an indicator parameter equal to 1 if customer i is served on route p and 0 otherwise. Let x_{pt} be a decision variable indicating how many times delivery pattern p is executed at period t and let d_{ipt} be the volume delivered to customer i using pattern p at period t and I_{it} be the inventory level at customer i at the end of period t. Then we formulate the mixed integer programming model as follows:

$$MIP: \quad \frac{1}{T} \min \sum_{p \in \mathcal{P}} \sum_{t=1}^{T} \theta_p x_{pt} \tag{54}$$

$$s.t. \ \sum_{i \in N} d_{ipt} a_{ip} \leq x_{pt} Q \quad \forall p \in \mathcal{P}, t \in \{1, \ldots, T\} \tag{55}$$

$$I_{i(t-1)} + \sum_{p \in \mathcal{P}} d_{ipt} a_{ip} = q_i + I_{it} \quad \forall i \in N, t \in \{1, \ldots, T\} \tag{56}$$

$$\sum_{p \in \mathcal{P}} d_{ipt} \leq C_i - I_{i(t-1)} \quad \forall i \in N, t \in \{1, \ldots, T\} \tag{57}$$

$$x_{pt} \in \mathbb{Z}^+ \tag{58}$$

$$d_{ipt} \geq 0 \quad \forall i \in N, p \in \mathcal{P}, t \in \{1, \ldots, T\} \tag{59}$$

$$I_{it} \geq 0 \quad \forall i \in N, t \in \{0, \ldots, T\}. \tag{60}$$

The objective function (54) expresses the minimization of the average transportation cost over the planning horizon. The constraints (55) guarantee that the total quantity delivered by each route in each period does not exceed the total capacity of the trucks serving that route on that period. The constraints (56) guarantee that no stockouts occur at any customer location during the planning horizon and also implicitly guarantee that a delivery to customer i on route p is positive only when customer i is visited on route p. The constraints (57) guarantee that no customer gets deliveries that will result in exceeding its storage capacity. Finally, the constraints (58), (59) and (60) are integrality requirements and sign restrictions.

A major disadvantage of the above mixed integer program is that every subset of the customer set corresponds to a delivery pattern hence the number of delivery patterns is exponential. Also, computing the cost of each pattern requires solving a traveling salesman problem. However, we can employ some techniques to reduce the number of patterns required to obtain the optimal solution or to approximate the optimal solution while using only a small subset of all the patterns.

As stated before, this mixed integer program provides a solution for the inventory routing problem considered in this chapter. Hence, the LP-relaxation of this MIP provides a lower bound on the optimal objective function value of the IRP.

4.3.2.2 Pattern Selection Linear Programming Model - A Lower Bound

Next, we review a method for computing a lower bound on the optimal objective function value of the IRP introduced by Song and Savelsbergh [40]. This method is similar to the pattern generation MIP discussed above with a modification in the pattern generation step. Here while generating the delivery routes, the volume to be delivered to each customer along the route is determined as well. Hence, the final model only selects the delivery routes.

Let P_j be a feasible pattern, $P_j = (d_{j1}, d_{j2}, \ldots, d_{jn})$ where d_{jn} denotes the amount delivered to customer n by pattern P_j. Hence P_j represents a delivery route that serves a subset of the customers, $\delta(P_j)$). It is clear that a feasible pattern satisfies $\sum_{i \in N} d_{ji} \leq Q$ and $0 \leq d_{ji} \leq C_i \quad \forall i \in N$. The cost of delivery pattern P_j, $\theta(P_j)$, is the optimal cost of the corresponding TSP. Let $\mathcal{P}$ be the set of all feasible delivery patterns. Let $\bar{x}_j$ be the variable denoting how many times pattern P_j is used and let U_i be the total demand of customer i during the planning period. The pattern selection LP below provides a lower bound on the average transportation cost over the planning horizon T:

$$PSLP: \quad \frac{1}{T} \min \sum_{j: P_j \in \mathcal{P}} \theta(P_j) \bar{x}_j \tag{61}$$

$$s.t. \sum_{j: P_j \in \mathcal{P}} d_{ji} \bar{x}_j \geq U_i \quad \forall i \in N \tag{62}$$

$$\bar{x}_j \geq 0. \tag{63}$$

The objective function (61) expresses the minimization of the average transportation cost over the planning horizon. The constraints (62) guarantee that each customer i is delivered at least its total demand over the planning horizon.

In this pattern generation model, the number of feasible patterns is considerably larger compared to the previous model. Given a delivery route with ordered customer (a pattern for the previous MIP model), any combination of delivery volumes to the customers along the route constitutes as a feasible pattern for this model as long as the truck capacity constraint is not violated.

Song and Savelsbergh [40] prove that only a subset of feasible patterns, *base patterns*, are sufficient to find the optimal solution to the pattern selection LP. Their definition of base patterns is as follows: "A feasible delivery pattern P is a base pattern if at most one customer, say k, receives a delivery quantity less than $\min\{C_k, Q\}$ in that case, the delivery quantity is $Q - \sum_{i \in \delta(P) \setminus \{k\}} C_i$".

4.3.2.3 Set Partitioning Model-An Upper Bound

Next, we develop an approach for finding a feasible distribution strategy for the supplier, hence obtaining an upper bound on the the optimal objective function value of the IRP. To obtain a feasible solution to the IRP, we will make a simplifying assumption that does not violate the feasibility of the distribution strategy.

An important phenomenon that affects the feasibility of a distribution plan is the interaction between the timing of the deliveries and the capacity restrictions. That is, the supplier can not deliver a volume to a customer that causes the storage capacity at the customer to be exceeded. The volume delivered to a customer at any time is bounded above by the minimum of the truck capacity and the remaining storage capacity at the customer location. However this is a complex constraint since the feasibility of a route with pre-specified delivery volumes depends on the remaining time-dependent capacities at the customer locations. Furthermore, if deliveries from multiple feasible routes overlap, the storage capacities at customer locations may be violated resulting in an infeasible distribution plan.

The complications due to the timing/capacity restrictions do not exist if we let any customer to be visited by only one vehicle route. In this special case of the IRP overlapping deliveries that may cause storage capacity violations are no longer possible and so the scheduling of deliveries is not an issue. In this case, when a group of customers is assigned to a route, this route is executed repeatedly over time to satisfy the demand of those customers. The major decision is to determine which customers should be grouped together on a route. In order to determine this, we first construct the feasible patterns as in the previous approaches, and then solve a set partitioning problem to determine which patterns are selected and hence which customers are grouped together.

Recall that in the first MIP model, we define a pattern as a set of customers to be visited in a given order and the volume delivered to each customer is determined by the MIP model. On the other hand, in the pattern selection LP model, the pre-constructed patterns not only specify the set of customers to be visited, but also determine the delivery volume to each customer. With the simplification that each customer is visited by only one route, the decision on which customers to be served on a route determines the optimal delivery volumes. That is, every time a route is executed, the truck visits every customer along the route and the delivery volumes to the customers are equal to an integer multiple of their consumption rate per period, $k \times q_i$. The maximum of such k values is bounded by the capacity of the truck and the storage capacities at the customer locations. As a result, we have a considerably smaller subset of patterns compared to the pattern selection LP model and prove that this is sufficient to find the optimal solution to the set partitioning model, which solves the special case of the IRP. The following lemma proves this result.

Lemma 10 *A feasible delivery pattern p is a base pattern if the ratio of the delivery quantity to the usage rate ($\frac{d_{ip}}{q_i}$) is the same for all the customers served by that pattern and is equal to the optimal "period" of the route. The optimal "period" of the route ($\frac{1}{freq_p}$) is equal to $\lfloor \min\{\min_{i\in\delta(p)}\{\frac{C_i}{q_i}\}, \frac{Q}{\sum_{i\in\delta(p)} q_i}\}\rfloor$. The base patterns are sufficient to* *find the optimal solution to the set partitioning model, which solves the special case of the IRP.*

Proof: The proof is fairly intuitive in the sense that route p is executed in every $\frac{1}{freq_p}$ period and visits all the customers belonging to that route. The period of a route is determined by $\frac{1}{freq_p} = \lfloor \min\{\min_{i\in\delta(p)}\{\frac{C_i}{q_i}\}, \frac{Q}{\sum_{i\in\delta(p)} q_i}\}\rfloor$. Note that a route is feasible if this ratio is positive. As long as the period of the route is fixed, it is optimal to deliver each customer a quantity large enough for satisfying the demand during the lead time (period) of the route. Thus, each customer along the route will receive an integer multiple of its consumption rate per period and this integer multiple value corresponds to the period of the route. □

We construct the feasible delivery patterns for the set partitioning model as described above. The cost of delivery pattern p, denoted by $\theta(p)$ is the optimal cost of the corresponding TSP. Let $\mathcal{P}$ be the set of all feasible base patterns and let $\hat{x}_p$ be a binary variable denoting whether pattern p is used or not. Let a_{ip} be an indicator parameter equal to 1 if customer i is served on route p and 0 otherwise. Then, the set partitioning problem below provides an upper bound on the average transportation cost over the planning horizon and also identifies a feasible distribution strategy for the supplier that visits each customer on exactly one delivery route.

$$SPM: \quad \min \sum_{p\in\mathcal{P}} freq_p \theta(p)\hat{x}_p \tag{64}$$

$$s.t. \sum_{p\in\mathcal{P}} a_{ip}\hat{x}_p = 1 \quad \forall i \in N \tag{65}$$

$$\hat{x}_p \in \{0,1\}. \tag{66}$$

The objective function (64) minimizes the average transportation cost. Note that, this model averages the costs over an infinite horizon. The reason for this is to obtain a repeatable distribution strategy and to avoid the complications resulting from the residual period's demands. The constraints (65) guarantee that every customer is visited by only one route.

Similar to the procedures discussed above, a major disadvantage of this model is that the number of delivery patterns may be exponential and computing the cost of each pattern requires solving a traveling salesman problem. However, as discussed above, the delivery volumes to the customers are not variables as in the MIP and the number of patterns is considerably less than the pattern selection LP model. Another disadvantage of the model is that the final problem to be solved is a set partitioning problem, which is known to be NP-Hard. However, this method is still useful especially in solving small instances of the problem and obtaining a feasible solution to the IRP.

4.3.2.4 Set Covering Model-An Approximation

Next, we present an approach that is inspired by both the pattern selection LP and the set partitioning model discussed above. Although this method neither generates a lower bound nor an upper bound on the optimal objective function of the IRP, it may provide a better approximation compared to the methods described above.

Recall the complexities generated by the timing/capacity restrictions. If we ignore timing considerations of deliveries, then we will considerably reduce the complexity of the problem. To overcome this complexity, if we do not care about generating a feasible distribution strategy, we can allow the deliveries from multiple routes to overlap and possibly violate the storage capacity restrictions at customer locations. With this simplifying assumption, we calculate the upper bound on the delivery volume of a customer as the minimum of the truck capacity and the storage capacity at the customer location.

In this model, we pre-construct the patterns with delivery volume information as in the pattern selection LP, and then solve a set covering problem to find an approximation for the optimal objective function value of the IRP. We refer this set covering model as SCM. Therefore, similar to the previous model, the number of feasible patterns may be exponential and we are required to solve a set covering problem.

The pattern generation procedure is similar to the "cutting stock problem", which is defined as the problem of minimizing the number of sheets of a particular width used to satisfy the demand of a set of items with smaller widths [6]. Here, the width of an item corresponds to the consumption rate of a customer and the width of the sheet corresponds to the capacity of the truck. Hence a feasible pattern p has the delivery volume information of the customers that are visited by that pattern and the delivery volume to a customer is an integer multiple of that customer's consumption rate. The period of the pattern is equal to the maximum of these integer multiple values, $\frac{1}{freq_p} = \max_{i \in N}\{\frac{d_{ip}}{q_i}\}$.

Lemma 11 *A feasible delivery pattern p is a base pattern if either $\sum_{i \in \delta(p)} d_{ip} + \min_{i \in \delta(p)} q_i > Q$ or $d_{ip} + q_i > C_i \;\; \forall i \in \delta(p)$. The base patterns are sufficient to find the optimal solution to the SCM.*

Proof: The proof is trivial. If both conditions are not satisfied, we can increase the delivery volume of at least one customer by its consumption rate per period. The modified pattern has the same cost with the original one and delivers a higher amount of product to at least one customer. Therefore, the modified pattern dominates the original one. It is easy to see that only non-dominated patterns are sufficient to find the optimal solution to the set covering model. $\square$

We construct the feasible non-dominated delivery patterns for the considered set covering model as described above. The cost of delivery pattern p,$\theta(p)$, is the optimal cost of the corresponding TSP. Let $\mathcal{P}$ be the set of all feasible base patterns and let $\tilde{x}_p$ be an integer variable denoting the number of times the pattern p is used and let

b_{ip} be a parameter equal to $freq_p \times \frac{d_{ip}}{q_i}$ for customer i and route p. The set covering model is as follows:

$$SCM: \quad \min \sum_{p \in \mathcal{P}} freq_p \theta(p) \tilde{x}_p \tag{67}$$

$$s.t. \sum_{p \in \mathcal{P}} b_{ip} \tilde{x}_p \geq 1 \quad \forall i \in N \tag{68}$$

$$\tilde{x}_p \in \mathbb{Z}^+. \tag{69}$$

The objective function (67) minimizes the average transportation cost. Note that, as above the costs are averaged over an infinite horizon. The constraints (68) guarantee that the demand of each customer is satisfied.

SCM does not provide an upper bound since its solution may result in deliveries that exceed customer capacities. SCM does not provide a lower bound either since it only allows a delivery volume that is an integer multiple of the customer's per period consumption rate. However, SCM does not assume that each customer is visited by only one route. Thus, the optimal solution to this method is likely to be a better approximation for the optimal distribution strategy of the original IRP than the solution of the SPM. It may also be preferable to the pattern selection LP model in the sense that this method will yield a distribution plan, whereas the pattern selection LP model provides only a lower bound value.

4.3.2.5 Comparison of the Models

Next, we discuss the practicality of the mixed integer program, MIP, the pattern selection LP, $PSLP$, the set partitioning model, SPM and the set covering model SCM.

First of all, all four methods require a pattern generation phase and depending on the method a pattern may specify a set of customers in a certain order and may include the delivery volume data of the customers. As stated before, the common handicap of all four methods is that the number of feasible patterns may be quite large and determination the cost of each feasible pattern requires to solve a TSP.

In MIP and SPM, for any given subset of customers, there exists at most one feasible pattern. In $PSLP$ and SCM, for any given customer subset, there may exist several feasible patterns. The exact numbers depend on the relationship between the parameters: truck capacity, storage capacity at customer locations and consumption rate of the customers. For an instance with 25 customers, the number of patterns for MIP and SPM is $2^{25} = 33,554,432$ provided that all of them are feasible. For an instance with 50 customer, the number becomes $2^{50} = 1.1259 \times 10^{15}$. Note that, for patterns having more than 3 customer, we need to solve a TSP to determine the cost values. Besides computational time concerns, such large numbers of patterns cause memory problems for large instances.

In order to use these methods for real-life size problems, we use the idea in Song and Savelsbergh [40], to limit the number of feasible patterns. Song and Savelsbergh [40] claim that it is common for some distribution systems such as in industrial gas distribution, to have a few customers along a delivery route. Using this practical knowledge, they limit the number of stops along a route to at most four customers and reduce the number of feasible patterns for $PSLP$ to an acceptable number. For example the number of feasible patterns for MIP and SPM with 25 and 50 customers reduces to at most 15,275 and 251,175, respectively. Moreover, limiting the number of stops along a route to four customers, significantly decreases the computational effort for solving the corresponding TSP to determine the costs of the patterns.

Next, we show a relationship between the optimal objective function values of the models. Let $z^*(N)$ be the optimal objective function value of the IRP, hence $z^*(N) = c(N)$. Let $z^*_{MIP}(N)$ be the optimal objective function value of the MIP model, which is also equal to $z^*(N)$. Also, let $z^*_{PSLP}(N)$, $z^*_{SPM}(N)$, and $z^*_{SCM}(N)$ be the optimal objective function values of models $PSLP$, SPM and SCM, respectively.

We prove the relationship between these values with the following lemma.

Lemma 12 *The relationship between the optimal objective function values of the models PSLP, SPM and SCM is as follows:* $z^*_{PSLP}(N) \leq z^*_{SCM}(N) \leq z^*_{SPM}(N)$

Proof: We first prove that $z^*_{PSLP}(N) \leq z^*_{SCM}(N)$. Note that every feasible pattern for the model SCM is also a feasible pattern for the model $PSLP$. Also, the parameter b_{ip} of model SCM is equal to $freq_p \times \frac{d_{ip}}{q_i}$ and the variable $\bar{x}_p$ of model $PSLP$ is equal to $freq_p T\tilde{x}_p$. The constraints (68) of SCM can be represented as

$$\sum_{p\in\mathcal{P}} b_{ip}\tilde{x}_p \geq 1 \quad \forall i \in N, \Rightarrow \sum_{p\in\mathcal{P}} d_{ip} freq_p \tilde{x}_p T \geq q_i T \quad \forall i \in N, \Rightarrow \sum_{p\in\mathcal{P}} d_{ip}\bar{x}_p \geq U_i \quad \forall i \in N,$$

which correspond to the constraints (62) of $PSLP$. Hence, the feasible region of $PSLP$ includes the feasible region of SCM. As for the objective functions, the objective function of SCM is equal to,

$$\min \sum_{p\in\mathcal{P}} freq_p \theta(p)\tilde{x}_p = \frac{1}{T} \min \sum_{p\in\mathcal{P}} \theta(p)\bar{x}_p,$$

which corresponds to the objective function of $PSLP$. As the objective functions of both model are equivalent and the fact that feasible region of $PSLP$ includes the feasible region of SCM, we conclude that,

$$z^*_{PSLP}(N) \leq z^*_{LP-SCM}(N) \leq z^*_{SCM}(N)$$

where $z^*_{LP-SCM}(N)$ represents the optimal objective function of the LP-relaxation of SCM.

For the second part of the proof ($z^*_{SCM}(N) \leq z^*_{SPM}(N)$), first note that every feasible pattern for the model SPM is also a feasible pattern for the model SCM. Also, the parameter a_{ip} of model SPM is equivalent to the parameter b_{ip} of model SCM since the delivery volumes of a feasible pattern of SPM is equal to $d_{ip} = \frac{q_i}{freq_p}$ for the customers along the route. Hence, $b_{ip} = freq_p \times \frac{d_{ip}}{q_i}$ is equal to 1 for the customers along the route and 0 otherwise, which is equivalent to the parameter a_{ip}'s values. Hence, the feasible region of SCM includes the feasible region of SPM. As the objective functions of both model are equivalent and the fact that the feasible region of SCM includes the feasible region of SPM, we conclude that,

$$z^*_{LP-SCM}(N) \leq z^*_{LP-SPM}(N) \quad and \quad z^*_{SCM}(N) \leq z^*_{SPM}(N)$$

where $z^*_{LP-SPM}(N)$ represents the optimal objective function of the LP-relaxation of SPM. $\square$

In addition to this result, we already know that $z^*_{PSLP}(N) \leq z^*(N) \leq z^*_{SPM}(N)$. Hence we formally conclude that the optimal solution value to the SCM model is a good approximation to the optimal solution value to the IRP.

4.4 Inventory Routing Game

In this section, we first define the cooperative game associated with the inventory routing problem, which is referred as the *inventory routing game*. Next, we propose methods to determine the cost-to-serve for the customers, equivalently we develop cost allocation methods for the cooperative game. We discuss the pros and cons of these allocation methods, in terms of practicality and theoretical properties. We also present examples to illustrate the behavior of the proposed methods.

4.4.1 Game Definition

We define the *inventory routing game (IRG)* as a cooperative game where the supplier has to serve the n customers that are the players in the game. The set of all the customers N is the grand coalition of the game and any subset of the customer $S \subset N$ makes up a coalition. The characteristic function $c(S)$ is the optimal average transportation cost of the IRP with a given set of customers S over the planning horizon and $c(N)$ is the total cost of the grand coalition. By defining the characteristic function in this manner, we implicity assume that there exist a sufficient number of trucks to serve any subset of the customers. The cost allocations for the IRG represent the cost-to-serve values for the customers.

We first prove that the IRG may have an empty core. That is, a budget balanced and stable cost allocation method may not always exist.

Lemma 13 *The core of the inventory routing game may be empty.*

Proof: Proof directly follows from the Traveling Salesman Game (TSG). The TSG easily reduces to an IRG with infinite truck and storage (at the customer locations) capacities if the cost matrix follows the triangular inequality assumption. As mentioned in the literature review section, there exist some instances of TSG with empty core, hence we conclude that the core of a IRG may be empty. We also present IRG instances with empty core later in the chapter. □

As stated before, the IRG is similar to the TSG and VRG discussed in the literature in the sense that in all three cooperative games we attempt to allocate the total transportation cost of the grand coalition among the customers. Also, in all three games computing the characteristic function value for a given coalition requires to solve an NP-Hard problem and calculating generic cost allocations such as the Shapley Value is impractical since it requires to consider an exponential number of subsets explicitly.

However, the IRG also has different characteristics compared to TSG and VRG. Most importantly, due to delivery volume decisions and the fact that IRP is a multi period problem, there exist practically an infinite number of feasible distribution patterns and so computing the exact characteristic function value is very challenging even for small instances of the problem.

We propose several cost allocation mechanisms for the inventory routing game. We classify our proposed methods into three categories: Proportional, Per-route based and Duality based cost allocation methods. In general proportional cost allocation methods are simple but easy to compute due to the fact that they do not take into account the effect of the interactions (or synergies) among the customers. Due to their simplicity, they are the most commonly used cost allocation methods in practice. On the other hand the other two methods consider the effect of customer synergies in some form and hence are harder to compute but provide more accurate cost-to-serve values.

One of the primary performance indicators of a cost allocation method is whether it is stable or not. If a cost allocation is stable than the allocated cost to a subset of customers cannot be greater than the stand alone cost of that subset, $c(S)$. However, stability is also a very restrictive condition and for most practical problems stable and budget balanced allocations do not exist. If the stability condition is not satisfied, than one can calculate the percentage deviation of the allocated cost of the subset from its stand alone cost. We call this percentage the *instability value* of a subset or coalition. Then, the instability value of a cost allocation method, is the maximum instability value over all the coalitions.

4.4.2 A Proportional Cost Allocation Method

In this subsection, we present a proportional cost allocation method. This method provides a quick cost allocation that takes into account some important factors such as customer distance to the depot, storage capacity and consumption rate but ignores some others such as synergies among the customers. Therefore, this method allows a quick estimation and more importantly does not require to compute (or approximate) the characteristic function of the game. Also, the procedure used by the gas distributor is similar to this method.

For the IRG, the parameters that define a customer's characteristics are distance to the depot and to the other customers, storage capacity, consumption rate and initial inventory level. Distance to the depot is the most relevant factor since the cost to be allocated among the customers is the total transportation cost of the problem and so the cost-to-serve value of a customer should be positively correlated to the customer's distance to the depot. Storage capacity and consumption rate are also significant since a customer's demand triggers a delivery in every $\lfloor \frac{C_i}{q_i} \rfloor$ period. Note that if $C_i > Q$ that customer's demand triggers a delivery in every $\lfloor \frac{Q}{q_i} \rfloor$ period. It is clear that the allocated cost of a customer should be negatively correlated to this value. The effect of initial inventory will fade over time therefore we do not consider this factor. The distances to other customers in a way reflect the geographical synergies between the customers, however it is far from a precise estimation. The reason for this is that we do not know which customers are actually grouped together on a route in advance and the synergies among customers also depend on the interactions among other factors such as storage capacities, consumption rates and the truck capacity. Therefore we do not include this parameter in our proposed method.

Let α_i be the cost allocated to customer i, then these values can be calculated with the following two step procedure:

PCAM:

Step 1 For each customer $i \in N$, compute $\beta_i = \frac{c_{i0}}{\lfloor \frac{\min\{C_i,Q\}}{q_i} \rfloor}$

Step 2 $\alpha_i = c(N) \times \frac{\beta_i}{\sum_{i \in N} \beta_i} \quad \forall i \in N.$

This procedure is very efficient and given an IRP instance provides a budget balanced allocation in seconds even for very large instances of the problem. Note that $2 \times \beta_i$ is equivalent to the cost of customer i if served individually. Therefore, this proportional cost allocation method allocates the total cost of the problem proportional to the individual costs of the customers.

4.4.3 A Per-route Based Cost Allocation Method

Next, we propose another cost allocation method, which allocates costs to the customers on a per-route basis. That is the cost of each delivery route is allocated to the customers it serves without considering the interactions among delivery routes. This cost allocation method is expected to perform better than the proportional cost allocation method discussed above since the cost allocation will be based on the optimal routes that actually compose the total cost and consider the intra-route synergies (synergies among the customers on a particular route).

The per-route based cost allocation method is suitable for the IRG in several aspects. First of all, the cost of every route is allocated to the customers that trigger the execution of that route so they are responsible for that cost. If the cost of each route is completely allocated among the customers it serves, then the resulting cost allocation will be a budget balance cost allocation. Per-route based cost allocations can be calculated fairly efficiently, especially when the delivery routes serve only a limited number of customers. For example, if every route has at most 4 customers, then the general IRG is reduced to several small games with 4 customers. In that case, even when the number of delivery routes is high, the computational effort required is relatively low since allocating cost for a 4-player game is not computationally difficult in general. Finally, in case of a modification in the IRP solution such as an update due to a new customer or demand, if the optimal solution of the IRP does not change considerably, it is relatively easy to update the cost allocations with this method.

Although any cost allocation method proposed for the TSG in the literature can be employed as a per-route based method for the IRG, we only consider the "moat packing" based method described in [19] since we can easily modify this allocation method to include delivery volume information. As mentioned in the literature review section, Faigle et al. [19] consider a TSG with a cost matrix satisfying the triangular inequality. By using the optimal value of the Held-Karp relaxation of the TSP, they obtain cost allocations from moat packings using LP duality. They prove that with their method, no coalition pays more than $\frac{4}{3}$ times its stand alone cost.

Below we present the basics of such a cost allocation method and our modifications to apply this method to the IRG.

4.4.3.1 A Moat Packing Based Cost Allocation for the TSG

In a TSP instance, the cities or the group of cities are surrounded by non-intersecting moats. To reach the cities outside of a particular moat, the vehicle has to cross the moat twice. Therefore, the customers outside any given moat are equally responsible for the vehicle to cross over that moat. As the maximum moat packings is a lower bound on the optimal cost of the TSP, it is reasonable to use these moat packings to allocate cost among the cities visited. In this cost allocation method, first a moat packing with maximum total width is determined and then the costs are distributed by allocating twice the widths (costs) of the moats among the cities on the outside of the moat.

Let $N \cup \{0\}$ be the set of cities of the TSP, "0" denoting the depot. Let S be a subset of the cities and $\bar{S}$ be the complementary of the set S and $\mathcal{M}$ be the set of all partitions $\{S, \bar{S}\}$. We assume that $0 \in \bar{S}$. Let $w_{S,\bar{S}}$ be the width of the moat surrounding the customer set S and let g_{ij} be the distance between cities i and j. The moat packing with maximum total width can be obtained by solving the linear program below:

$$MP: \quad \max \sum_{S,\bar{S} \in \mathcal{M}} w_{S,\bar{S}} \tag{70}$$

$$s.t. \sum_{i \in S, j \in \bar{S}} w_{S,\bar{S}} \leq g_{ij} \quad \forall i, j \in N \cup \{0\} \tag{71}$$

$$w_{S,\bar{S}} \geq 0 \quad \forall \{S, \bar{S}\} \in \mathcal{M}. \tag{72}$$

The formulation above always has an optimal moat packing that has a *nested* structure (that is no two moats intersect). After finding a nested moat packing with maximum total width, the allocated costs is determined by distributing twice the width of any moat among the cities on the outside of the moat in an arbitrary way.

$$\alpha_i = 2 \times \sum_{S,\bar{S} \in \mathcal{M} i \in S, 0 \in \bar{S}} \frac{w^*_{S,\bar{S}}}{\mid S \mid}$$

It is proved that in the resulting cost allocation no coalition pays more than its stand alone cost, which is equal to the optimal cost of the corresponding TSP. However, the maximum total width of moat packing may be less than the optimal cost of the associated TSP, hence the cost allocation method may not be budget balanced. Although we can scale up the allocated costs to achieve a budget balanced cost allocation, in that case, the resulting cost allocation may not be stable. Also, the width of the moats can be allocated in a more structured way. For example, $\frac{1}{|S|}$ may be replaced by the weights λ^i_S. Then, $\alpha_i = 2 \times \sum \lambda^i_S w^*_{S,\bar{S}}$ will be the cost allocated to city i.

Another issue with this method is that MP may have exponentially many variables. Therefore, Faigle et al. [19] propose to solve the dual problem in polynomial time. They also claim that an optimal moat packing can be determined by computing the dual variables of the dual problem, which correspond to the moat width variables of the primal problem, in polynomial time. We believe that this claim is not rigorous since this procedure does not guarantee a nested moat packing. We propose a procedure, although not a polynomial one due to exponentially many variables, to obtain a nested moat packing from a non-nested one in Appendix B. We believe that the question of if a nested moat-packing can be obtained in polynomial time remains open.

Furthermore, for a TSG instance with a non-empty core, the moat packing based method may not be able to find an cost allocation in the core (see [19] for an example).

4.4.3.2 A Moat Packing Cost Allocation for the IRG

In order to devise a moat-packing based cost allocation method for the IRG, we modify the weights of the cities (customers) to reflect the delivery volumes. In a

TSG, uniform weights are appropriate since the cities outside the moat are equally responsible for the vehicle to cross over that moat. However, in the IRG context, the customers may have different levels of influence on the execution of a route, so they may not be equally responsible for the TSP tour serving them.

Let the *ideal delivery volume* to a customer be the minimum of the truck capacity and the storage capacity at the customer location ($\min\{Q, C_i\}$). As a general principle, the customer that receives an ideal delivery volume from a route should be allocated a higher cost than a customer that receives the residual amount in the vehicle, if both customers happen to be at the same location. Let r_i be the ratio of the actual volume delivered to a customer location by the route and the ideal delivery volume. Then, we modify the weights for the customers as follows:

$$\lambda_S^i = \frac{r_i}{\sum_{i \in S} r_i}.$$

Using these modified weights in distributing the moat costs, we compute the cost allocated to each customer on a given delivery route. Note that this method considers each route individually, hence the weights for the customer are computed for every delivery route. The moat packing cost allocation method for the IRG is as follows:

MPCAM:

Step 1 From the solution of IRP, identify the optimal delivery routes. Let K be the set of optimal delivery routes. For each route $k \in K$, execute steps 2-5.

Step 2 For each customer i on route k, calculate $r_i(k)$.

Step 3 Solve the dual MP and determine an optimal nested moat packing and $w(k)$ values using the procedure discussed above.

Step 4 Calculate the allocated cost of a customer along the route using the following formulas:

$$\lambda_S^i(k) = \frac{r_i(k)}{\sum_{i \in S} r_i(k)}, \quad \alpha_i(k) = 2 \times \sum \lambda_S^i(k) w^*_{S,\bar{S}}(k)$$

Step 5 If the total allocated cost to the customers is less than the cost of the route, scale up the cost allocation values to obtain a budget balanced cost allocation.

Step 6 After computing the allocated cost of each customer for every delivery route, compute the final cost allocation of customers α_i by summing up the cost allocations from individual routes, $\alpha_i = \sum_{k \in K} \alpha_i(k) \;\; \forall i \in N$.

This per-route based cost allocation method will perform better in terms of stability compared to the proportional cost allocation method because within each route it considers the stability of the allocation. Therefore, such a cost allocation method works best if the inter-route synergies are minimal. Furthermore since the optimal delivery routes represent the subset of customers with the highest synergies and given that this method is likely to achieve the intra-route stability for each route, we expect that it will perform better than the proportional cost allocation method.

Although this per-route based cost allocation method has many advantages in terms of stability and computational efficiency, it also has some limitations. First of all, this method requires not only the optimal cost of the IRP, but also the optimal delivery routes of the problem. Second, if the optimal IRP solution is not available, this method allocates the costs to the customer based on the given set of delivery routes, which creates a bias on the performance of this cost allocation method.

4.4.4 Generic Cost Allocation Methods

There exist a set of generic solution methods proposed in the literature for the cost allocation problem of any cooperative game. One of these methods involves finding the *Nucleolus*, which is defined as the cost allocation that lexicographically maximizes the minimal gain for the coalitions. Another well-known cost allocation method distributes the costs with respect to the *Shapley Value*, which is based on the marginal cost of each player to each coalition when forming the grand coalition. Nucleolus is guaranteed to be in the core if the core is non-empty, however, Shapley Value is not.

Both of these methods require the computation of the characteristic function of each coalition explicitly. Therefore, any approach that computes the actual cost allocations obtained by the nucleolus or the Shapley value is expected to have an exponential running time due to exponential number of coalitions. Also, note that for each coalition we require to solve a NP-hard problem if we were to compute the nucleolus or the Shapley value exactly. Due to these limitation, we do not explore these methods further in this chapter.

4.4.5 Duality Based Cost Allocation Methods

In the literature, for several games based on combinatorial optimization problems, the relationship between its core and the dual of the LP-relaxation of an Integer Programmimg formulation of the problem is well established. For various games such as facility location, set partitioning, packing, covering, it is proven that the absence of an integrality gap for the LP-relaxation is a necessary and sufficient condition for the non-emptiness of the core.

In our context, we have a model (MIP) that exactly solves the underlying optimization problem, IRP, however solving this model to optimality may not be possible for a real-life size instance. Also, the LP-relaxation for the MIP model may not provide a tight lower bound on the optimal objective function of the IRP, hence it may not be a good basis for a duality based cost allocation method.

We have also developed three different models that yield upper and lower bounds and an approximation to the optimal objective function value of the IRP. Therefore, instead of using just one LP-relaxation, we consider the four different LP's: LP-relaxation of the mixed integer programming model ($MIP - LP$), pattern selection

LP ($PSLP$), LP-relaxation of the set partitioning model ($SPM - LP$), and finally LP relaxation of the set covering model ($SCM - LP$). Although none of the dual problems corresponding to these LP's will give the desired cost allocation exactly, we modify their solutions to construct a cost allocation for the IRG.

An important point in using the dual of the LP-relaxations to obtain a cost allocation is that the total cost that can be recovered is bounded by the optimal objective function value of the LP-relaxation. If the total transportation cost of the optimal distribution strategy is greater than the cost recovered by the cost allocation methods based on the dual problems, we then simply scale up the cost allocations.

When we scale up (or down) such cost allocations to get a budget balanced cost allocation for the IRG, the resulting cost allocations may not be stable. Hence, we desire LP-relaxations that are tighter approximations. We claim that $MIP-LP$ will provide a looser lower bound on the optimal objective function of the IRP, compared to the other methods. Hence, a cost allocation obtained by the dual of $MIP - LP$ will either recover a smaller fraction of the total budget or have a higher violation of stability compared to the other duality based methods.

It is possible that cost allocations from any of the duality based methods will be instable, when there exists a cost allocation in the core for the IRG. However, it may also be the case that the particular IRG has an empty core and no cost allocation method could do any better.

The duality based cost allocation methods have several advantages over the other cost allocation methods discussed before. First of all, duality based methods do not require any IRP solution information, such as optimal delivery routes. These methods have a system wide perspective and take both intra-route and inter-route synergies into account, hence cost allocations from duality based method are more likely to perform better in terms of stability. Also, duality based cost allocation methods are not biased by delivery route information as the per-route allocation method and do not penalize a customer for being on a high cost delivery route. That is, suppose we have two identical customers at the same location and one of them is served by a cost efficient route and the other one is not. Since these customers are identical from every aspect, the duality based cost allocation methods will compute the same cost-to-serve value for both customers. On the other hand, the per-route based cost allocation method will treat them differently as that method allocates cost on a per-route basis. Finally, duality based methods are also computationally efficient, since they only require to solve an LP and possibly perform a scaling up operation to obtain a budget balanced cost allocation.

Next, we construct the dual LP's corresponding to each solution method we proposed in Subsection 4.3.2 and compute cost allocations for the IRG based on these dual LP's.

4.4.5.1 A Cost Allocation Method Based on Dual MIP-LP

For the MIP, let u_{pt} be the dual variables associated with constraints (55), y_{it} be the dual variables associated with constraints (56), and v_{it} be the dual variables associated with constraints (57). Then the dual of the LP-relaxation of the MIP is as follows:

$$DMIP - LP: \quad \max \sum_{i \in N} \sum_{t \in T} q_i y_{it} + \sum_{i \in N} \sum_{t \in T} C_i v_{it} \tag{73}$$

$$s.t. \; -Qu_{pt} \leq \frac{\theta(p)}{T} \;\; \forall p \in \mathcal{P}, t \in \{1, \ldots, T\} \tag{74}$$

$$v_{i(t+1)} - y_{it} + y_{i(t+1)} \leq 0 \;\; \forall i \in I, t \in \{1, \ldots, (T-1)\} \tag{75}$$

$$-y_{iT} \leq 0 \;\; \forall i \in I \tag{76}$$

$$a_{ip} u_{pt} + a_{ip} y_{it} + v_{it} \leq 0 \;\; \forall i \in I, p \in \mathcal{P}, t \in \{1, \ldots, T\} \tag{77}$$

$$u_{pt} \leq 0, \quad v_{it} \leq 0, \quad y_{it} \;\; u.r.s. \tag{78}$$

Let $z^*_{DMIP-LP}(N)$ be the optimal objective function of $DMIP - LP$ and (u^*, y^*, v^*) be the corresponding optimal solution that can be computed in polynomial time.

Let α_i be the allocated cost to customer $i \in N$. Then the $DMIP - LP$ based cost allocation method for the IRG is as follows:

DMIP-LP:

Step 1 Solve $DMIP - LP$ and determine (u^*, y^*, v^*).

Step 2 Calculate the allocated costs of the customers using the following formula:

$$\alpha_i = \frac{z^*(N)}{z^*_{DMIP-LP}(N)} \sum_{t \in T} q_i y^*_{it} \quad \forall i \in N.$$

Note that multiplying by $\frac{z^*(N)}{z^*_{DMIP-LP}(N)}$ corresponds to the scaling-up step to obtain a budget balanced cost allocation for the IRG.

4.4.5.2 A Cost Allocation Method Based on Dual PSLP

For the $PSLP$, let y_i be the dual variables associated with constraints (62). Then the dual of the pattern selection LP is as follows:

$$DPSLP: \quad \max \sum_{i \in N} U_i y_i \tag{79}$$

$$s.t. \sum_{i \in N} d_{ji} y_i \leq \frac{\theta(P_j)}{T} \quad \forall P_j \in \mathcal{P} \tag{80}$$

$$y_i \geq 0. \tag{81}$$

Let $z^*_{DPSLP}(N)$ be the optimal objective function of the $DPSLP$ and y^* be the corresponding optimal solution. Let α_i be the allocated cost to customer $i \in N$. The $DPSLP$ based cost allocation method for the IRG is as follows:

DPSLP:

Step 1 Solve $DPSLP$ and determine y^*.

Step 2 Calculate the allocated costs of the customers using the following formula:

$$\alpha_i = \frac{z^*(N)}{z^*_{DPSLP}(N)} U_i y^*_i \quad \forall i \in N$$

4.4.5.3 Cost Allocation Method Based on Dual SPM-LP

Next, for the $SPM-LP$, let y_i be the dual variables associated with constraints (65) and let u_p be the dual variables associated with constraints (66), then the dual of the LP-relaxation of SPM is as follows:

$$DSPM-LP: \quad \max \sum_{i \in N} y_i + \sum_{p \in \mathcal{P}} u_p \tag{82}$$

$$s.t. \sum_{i \in N} a_{ip} y_i + u_p \leq \theta(p) \quad \forall p \in \mathcal{P} \tag{83}$$

$$y_i \quad u.r.s., \quad u_p \leq 0. \tag{84}$$

Let $z^*_{DSPM-LP}(N)$ be the optimal objective function of $DSPM-LP$ and (y^*, u^*) be the corresponding optimal solution that can be computed by solving $DSPM-LP$ in polynomial time. Then the $DSPM-LP$ based cost allocation method for the IRG is as follows:

DSPM-LP:

Step 1 Solve $DSPM-LP$ and determine y^*.

Step 2 Calculate the allocated costs of the customers using the following formula:

$$\alpha_i = \frac{z^*(N)}{z^*_{DSPM-LP}(N)} y^*_i \quad \forall i \in N.$$

4.4.5.4 Cost Allocation Method Based on Dual SCM-LP

Finally, for the $SCM-LP$, let y_i be the dual variables associated with constraints (68). Then the dual of the LP-relaxation of SCM is as follows:

$$DSCM-LP: \quad \max \sum_{i \in N} y_i \tag{85}$$

$$s.t. \sum_{i \in N} b_{ip} y_i \leq \theta(p) \quad \forall p \in \mathcal{P} \tag{86}$$

$$y_i \geq 0. \tag{87}$$

Let $z^*_{DSCM-LP}(N)$ be the optimal objective function of $DSCM - LP$ and y^* be the corresponding optimal solution. Then the $DSCM-LP$ based cost allocation method for the IRG is as follows:

DSCM-LP:

Step 1 Solve $DSCM - LP$ and determine y^*.

Step 2 Calculate the allocated costs of the customers using the following formula:

$$\alpha_i = \frac{z^*(N)}{z^*_{DSCM-LP}(N)} y^*_i \quad \forall i \in N.$$

4.4.6 Illustrative Examples

In this subsection, we demonstrate the performance of the methods we proposed for the IRG on small instances. In general, we show the limitations of the proportional and per-route based cost allocation models compared to the duality based ones. However note that these methods' computational efficiencies cannot be apprehended on these small instances. .

In the first example, there are three customers A, B and C. The location of these customers as well as the parameters of the instance are presented in Figure 14. The optimal delivery routes for this instance serve customer A and B together and customer C alone. As can be seen from Figure 14, there exist only intra-route synergies between customers A and B.

The columns of Table 8 present the allocations based on the different methods: proportional cost allocation method PRO, per-route based cost allocation method $MOAT$, and duality based methods $DMIP - LP$, $DPSLP$, $DSPM - LP$, and $DSCM - LP$. The rows represent the corresponding allocated costs to the customers with each method. The last row represents the total cost allocated with the respective cost allocation methods, which is basically the optimal cost of the IRP for this instance.

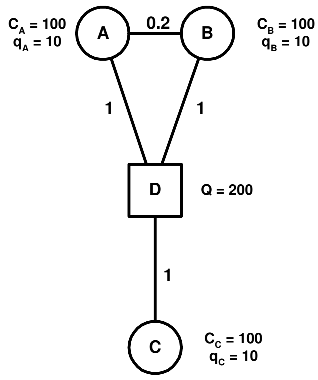

Figure 14: An instance with an empty core.

Table 8: Cost allocations for Example 1.

	PRO	MOAT	DMIP-LP	DPSLP	DSPM-LP	DSCM-LP
A	1.4	1.1	1.4	1.1	1.1	1.1
B	1.4	1.1	1.4	1.1	1.1	1.1
C	1.4	2	1.4	2	2	2
Total	4.2	4.2	4.2	4.2	4.2	4.2

Note that all allocations except for the PRO and $DMIP-LP$ realize the positive intra-route synergy among customers A and B and allocate a smaller cost to these than to customer C, hence are stable. The reason that $DMIP - LP$ provides such an allocation is that LP-relaxation of the MIP yields a looser lower bound compared to the other LP-relaxations. As stated before, the quality of a duality based cost allocation is directly related to how close the optimal objective function of the dual problem to the optimal objective function of the IRP. Therefore, $DMIP - LP$ does not perform as well as the other duality based methods.

In our second example, presented in Figure 15, there are three customers, A, B and C, which are identical in every sense. All the parameters of the instance are included in Figure 15. Hence, the desired cost allocation should allocate the same cost to all three customers. The optimal distribution strategy is to visit any two customers together, and the other customer alone every period. Also, note that this example is an instance with an empty core.

From Table 9, we observe that the proportional cost allocation method and the duality based cost allocation methods treat all there customers equally as expected. However, the cost allocation of the per-route based method depends on the distribution strategy used over the planning horizon. For example, if the same two customers are visited together every period then their cost allocations will be 1.866 and the cost allocation of the third customer will be 2. If we alternate the customers visited together every period, then the cost allocated to all three customers will be 1.911. This is a disadvantage of the per-route based cost allocation method, since it is based on the delivery routes used rather than the actual dynamics of the instance.

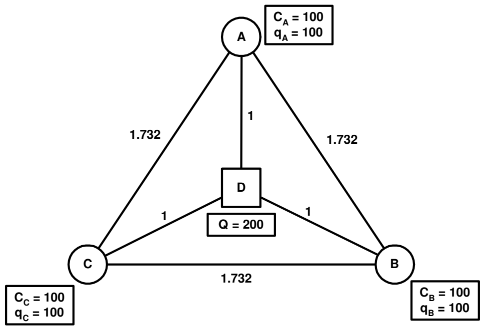

Figure 15: An instance with an empty core.

Table 9: Cost allocations for Example 2.

	PRO	MOAT	DMIP-LP	DPSLP	DSPM-LP	DSCM-LP
A	1.911	NA	1.911	1.911	1.911	1.911
B	1.911	NA	1.911	1.911	1.911	1.911
C	1.911	NA	1.911	1.911	1.911	1.911
Total	5.732	5.732	5.732	5.732	5.732	5.732

The third example, Figure 16, is an empty core instance with a single route optimal distribution strategy. Therefore, any cost allocation will have some instability, hence our objective is to find the one with the lowest instability. There exist both intra-route and inter-route synergies between customers, hence we expect that the duality based methods to perform better than the proportional and per-route based methods.

For example, customer A should be allocated a lower cost compared to B and C due to its synergies with both B and C.

As can be seen from Table 10, the proportional, per-route based, and $DMIP-LP$ cost allocations are unable to asses the advantage of customer A, and treat all three customers equally. On the other hand, the other duality based cost allocation methods are able to differentiate the customers and allocate the cost accordingly. The cost allocation method that works best on this instance is $DSPM-LP$ since it provides a cost allocation with the lowest instability value (2.75%).

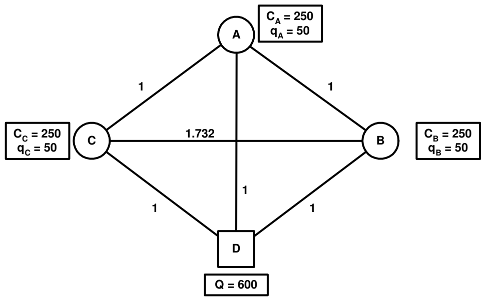

Figure 16: An empty core instance with a single route optimal solution.

Table 10: Cost allocations for Example 3.

	PRO	MOAT	DMIP-LP	DPSLP	DSPM-LP	DSCM-LP
A	0.333	0.333	0.333	0.263	0.233	0.263
B	0.333	0.333	0.333	0.368	0.383	0.368
C	0.333	0.333	0.333	0.368	0.383	0.368
Total	1.000	1.000	1.000	1.000	1.000	1.000

The fourth example, Figure 17, is again a non-empty core instance with a single route optimal distribution strategy. This example shows that even though there is only one route whose cost to be allocated, the per-route cost allocation can still perform quite unsatisfactorily. In this example, customer A will receive a low cost if it is served alone, however by the per-route based cost allocation it is assigned a cost more than five times its stand alone cost. By changing the values of the parameters, we can increase this instability value to infinity, which proves that the per-route based cost allocations that do not consider inter-route synergies may not work well even on instances with single route optimal solutions.

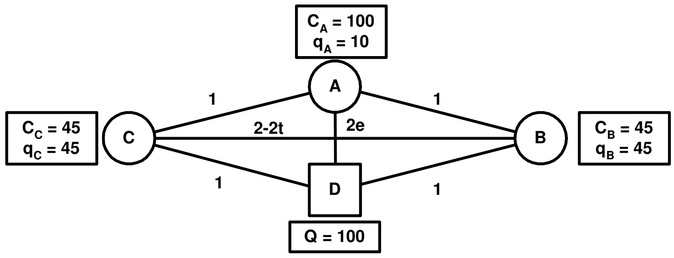

Figure 17: Another non-empty core instance with single route optimal solution.

Table 11: Cost allocations for Example 4.

	PRO	MOAT	DMIP-LP	DPSLP	DSPM-LP	DSCM-LP
A	0.040	0.205	0.087	0.018	0.039	0.039
B	1.980	1.898	1.956	1.991	1.981	1.981
C	1.980	1.898	1.956	1.991	1.981	1.981
Total	4.000	4.000	4.000	4.000	4.000	4.000

The fifth example, Figure 18 is an empty core instance where there are 11 identical customers at the same location and a vehicle is able to serve only 10 customers. The total transportation cost per period is 4 and since all customers are identical they should be all allocated $\frac{4}{11}$. Therefore, the total cost allocated to any coalition with 10 customers is $\frac{40}{11}$ and is greater than the stand alone cost of the coalition, which is equal to 2. The instability value of the cost allocation approximates to 100% as the number of customers goes to infinity.

From these results, we observe that the proportional and per-route based cost allocations have some limitations depending on the instance characteristics, especially when the synergies between customers are high. Duality based cost allocation methods, expect $DMIP - LP$, perform quite satisfactory in all examples presented above.

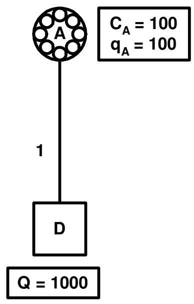

Figure 18: An empty core instance with 100% instability.

4.5 Computational Study

We have carried out a computational study on random instances to evaluate the performance of the methods we have discussed in terms of computational efficiency the stability of the allocations they generate. Over a 1,000 × 1,000 square, we generated 50 different instance with 25 and 50 customers. We set the truck capacity to 1000 in all instances. We generated two types of customers with high or low storage capacities. The ratio $\frac{Q}{C_i}$ for customer i falls within [0.5, 0.75] or [1.25, 1.5] for high or low storage capacity customers, respectively. Similarly, the ratio of the consumption rate to the storage capacity is either within [0.5, 0.75] or [1.25, 1.5] specifying that the customer has a relatively low or high consumption rate. As some of our proposed methods such as SPM or SCM require that the consumption rate of the customers to be less than both the truck capacity and the storage capacity, we allow multiple deliveries in a period to satisfy this assumption. That is, we divide a period into subperiods until the consumption rate of all the customers in a subperiod are less than both the truck capacity and the storage capacities. The planning horizon in all the instances is 100 periods and each period may include several subperiods depending on the parameters.

Two other parameters, clusters and ratio of cluster points to total points, allow us to represent the dense regions with many supply or demand points such as metropolitan areas and to control how many of the total points fall within a cluster. Clusters are uniformly distributed in the square and each point in a cluster is generated by using a standard Normal distribution. Remaining points are distributed uniformly across the map. In the instances with 25 customers, there are 3 clusters and in the instances with 50 customers, there are 4 clusters and in both cases 45% of the customer locations fall within the clusters.

We test the performance of the proposed cost allocation methods in both solution quality and computational time. The solution quality refers to the maximum percentage instability of a cost allocation method. However, calculating this value is computationally challenging. To calculate the exact percentage instability of a cost allocation method, we first need to calculate the characteristic function values for all coalitions of the collaboration, which requires solving an IRP for each coalition. Next, unless an implicit method is devised we have to calculate the percent instability value for each coalition. To this end, we test the stability of coalitions of size less than or equal to 4. Even this task requires evaluating 251,175 coalitions for the instances with 50 customers. Further, it is reasonable to assume that in practical settings due to limited rationality and information sharing of the customers it is unlikely for larger sub-coalitions to form.

Since IRP is NP-hard, instead of $z^*_{MIP}(N)$, we use approximations for the characteristic function value of the grand coalition. Recall that SPM gives an upper bound on the optimal objective function value of the IRP and hence $z^*_{SPM}(N)$ is a candidate for such an approximation. Another option is to use the optimal objective function value of $SCM - LP$, $z^*_{LP-SCM}(N)$, as this value can be computed efficiently by solving a linear program. Although we know that $z^*_{MIP}(N) \leq z^*_{SPM}(N)$ and $z^*_{LP-SCM}(N) \leq z^*_{SPM}(N)$, the relationship between $z^*_{LP-SCM}(N)$ and $z^*_{MIP}(N)$ is not clear. Hence, we cannot claim that any of these approximations gives a more accurate representation of the exact characteristic function value.

For the characteristic function values of the coalitions, instead of $z^*_{MIP}(S)$ we use the upper bound value, $z^*_{SPM}(S)$. However, note that using this value will underestimate the maximum percentage instability value as well as the number of instable subsets. Also using $z^*_{SPM}(S)$ will favor the cost allocation method based on $DSPM$ compared to the other methods.

The per-route based cost allocation method requires the delivery routes to allocate the costs among the customers. Because we cannot solve the IRP to optimality, we use the selected routes (patterns) of the $PSLP$ method as an input for the $MOAT$ cost allocation method. The drawback of using these routes is that the $MOAT$ will yield a biased cost allocation that is similar to the cost allocation of $DPSLP$ method. Also, we do not present the statistics of $DMIP-LP$ allocation as this allocation method's performance is significantly worse compared to the other dual based methods and sometimes even worse than the proportional cost allocation method as demonstrated in Subsection 4.4.6.

Table 12 summarizes the results of the experiments in which we use $z^*_{LP-SCM}(N)$ as an approximation to the optimal total cost of the IRP, $c(N)$. The columns represent the cost allocation methods as before. The first four rows summarize the performance of the methods on the instance with 25 customers. The first row presents the average number of instable subsets (up to size 4) with each cost allocation method. The row "AVE" presents the average of average percent instability of cost allocation methods over all the subsets whereas the next row "MAX" presents the average of maximum percent instability of cost allocation methods over all the subsets. The next row, "MAX MAX" shows the maximum of maximum percent instability of cost allocation methods over all the subsets and instances. The next four rows present the same performance indices for the instances with 50 customers. Similarly, Table 13 summarizes the performance of the cost allocation methods when we use $z^*_{SPM}(N)$ as an approximation to the optimal total cost of the IRP.

Our computational experiments show that duality-based cost allocation methods

Table 12: Instability values of the proposed methods when the IRP optimal cost is approximated by $z^*_{LP-SCM}(N)$

25 Customers	PRO	MOAT	DPSLP	DSPM-LP	DSCM-LP
# of subsets	2071.08	1318.52	1319.04	0.00	82.92
AVE	5.86	3.96	3.90	0.00	0.00
MAX	32.66	16.69	14.16	0.00	0.00
MAX MAX	51.26	36.44	21.93	0.00	0.00
50 Customers	**PRO**	**MOAT**	**DPSLP**	**DSPM-LP**	**DSCM-LP**
# of subsets	16679.00	9236.44	9433.04	0.00	336.64
AVE	5.32	2.69	2.65	0.00	0.00
MAX	32.91	13.52	11.82	0.00	0.00
MAX MAX	49.64	23.21	22.02	0.00	0.00

Table 13: Instability values of the proposed methods when the IRP optimal cost is approximated by $z^*_{SPM}(N)$

25 Customers	PRO	MOAT	DPSLP	DSPM-LP	DSCM-LP
# of subsets	2942.68	2209.36	2189.92	341.08	859.68
AVE	6.95	4.99	4.95	0.12	1.70
MAX	36.90	20.46	17.85	0.15	3.24
MAX MAX	53.91	42.46	26.51	1.10	5.88
50 Customers	**PRO**	**MOAT**	**DPSLP**	**DSPM-LP**	**DSCM-LP**
# of subsets	25979.64	16305.68	16557.04	2124.44	4694.72
AVE	5.50	3.32	3.33	0.14	1.23
MAX	36.08	16.25	14.50	0.17	2.40
MAX MAX	52.38	27.76	24.25	0.57	5.06

$DSPM-LP$ and $DSCM-LP$ perform significantly better than the other cost allocation methods. Even their maximum of maximum percentage instability values are less than 6%. As expected the proportional cost allocation method has a high number of instable subsets and extremely high maximum instability values. The average of maximum percentage instability value is around 35% and the maximum of this value is more than 50%. As we mentioned before, per-route based cost allocation method, $MOAT$, mimics $DPSLP$ since it uses the solution of the latter method as an input. Both methods perform better compared to the proportional cost allocation method, however, are significantly outperformed by $DSPM-LP$ and $DSCM-LP$.

Table 14 summarizes the computational time of the proposed cost allocation methods. The computational times of the proportional and per-route based cost allocation methods are negligible and hence are not presented in Table 14. As can be seen from Table 14, all the proposed cost allocation methods complete within seconds. Hence we conclude that our methods are not only effective but also computationally efficient.

Table 14: Computational times of the proposed methods in CPU seconds

25 Customers	**DPSLP**	**DSPM-LP**	**DSCM-LP**
Ave	0.32	0.04	0.76
Max	1.00	1.00	4.00
50 Customers	**DPSLP**	**DSPM-LP**	**DSCM-LP**
Ave	0.40	2.84	36.20
Max	1.00	6.00	203.00

4.6 Conclusion

In a vendor managed inventory setting, assessing the cost-to-serve information of the customers is of value when charging prices to the customers, prioritizing deliveries, marketing to new customers, or revisiting the routing/volume decisions. We propose several cost allocation methods to determine the cost-to-serve for the customers. We show that empirically our proposed methods perform significantly better than proportional allocation methods used in practice as these methods ignore the synergies among the customers. We also demonstrate that the proposed methods are computational efficient and can be implemented for solving real-life problems.

APPENDIX A

PROOF OF THEOREM 3

Theorem 3 *When there are three or more carriers with non-identical cost structures, then MCLCP is NP-hard.*

Proof: By a reduction from 3-SAT. In 3-SAT, we have a set of boolean variables $\mathcal{B} = \{B_1, \ldots, B_n\}$ and a set of clauses $\mathcal{U} = \{CLA_1, \ldots, CLA_m\}$. Each clause has exactly three literals, where a literal corresponds to a boolean variable B_i or its complement $\bar{B}_i$. A clause is satisfied when it contains at least one true literal. The question is whether every clause can be satisfied by setting each boolean variable to either true or false.

We give a polynomial reduction from 3-SAT to $MCLCP$ with three or more carriers and non-identical cost structures ($3^+ - MCLCP$).

Each clause is converted to an arc in such a way that these arcs form a path, the so-called *clause path* (see Figure 19). For each boolean variable B_i, we construct a

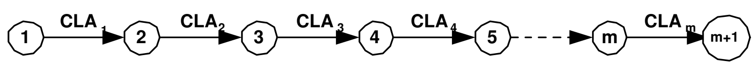

Figure 19: The clause path.

network with an upper path corresponding to the variable itself, and a lower path corresponding to its complement. An example is depicted in Figure 20, where the upper path is shown in solid lines and the lower path is shown in dashed lines. These paths intersect the clause path whenever the variable appears in the clause. In the example, the upper path intersects with the clause path twice, in arcs (2,3) and (4,5), indicating that variable B_i is included in both CLA_2 and CLA_4. Similarly, the lower path intersects with the clause path once, in arc (3,4), indicating that $\bar{B}_i$ is only included in CLA_3. The network is completed with two more arcs: a *shared arc* and a *returning arc*; also shown in Figure 20. The lane set L consists of the arcs in the clause path and the shared arcs in each of the networks associated with the boolean variables. We set the r_{ij} value for each clause arc and for each shared arc to one. The latter guarantees that there is at least a unit flow on either the upper or the lower path of every boolean variable.

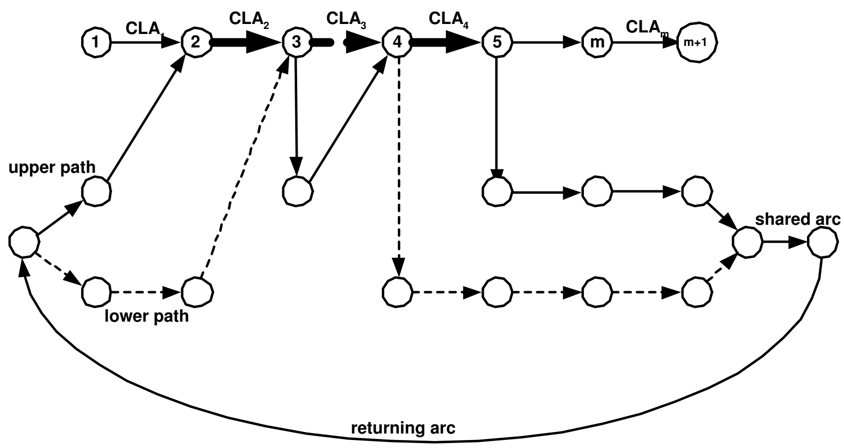

Figure 20: Network corresponding to a boolean variable.

Next, we create a carrier for each boolean variable, in the sense that carrier can only cover arcs in the network associated with the boolean variable. This can be done by setting the cost of all other arcs to infinity for that carrier. This guarantees that a flow in the network associated with a boolean variable stays on the lower or upper path. For the arcs in the network associated with a boolean variable, we set the costs to zero except for the *shared arc* which gets a cost of one. Repositioning costs are assumed to be equal to the original lane costs for every carrier.

Note that each clause arc appears in exactly three boolean variable paths (either an upper or a lower path). A clause is satisfied if and only if the corresponding clause arc is covered by at least one of the three carriers associated with the boolean variables. Note that by setting r_{ij} equal to one for all the clause arcs, we ensure that each clause in $\mathcal{U}$ is satisfied. The constructed instance of $3^+ - MCLCP$ is always feasible since by sending a unit flow on all upper and lower paths corresponding to the boolean variables we will satisfy every clause. In that case, the cost for each carrier is equal to two, since the cost of the shared arc is equal to 1, and thus the total cost of the carriers is equal to $2 \times \mid n \mid$. Hence, an instance of 3-SAT is feasible if and only if the corresponding $3^+ - MCLCP$ instance has an optimal solution with a value of $\mid n \mid$. Any feasible solution with cost strictly greater than $\mid n \mid$ will have at least one boolean variable with a flow of one on both the upper and lower path, which implies that the 3-SAT instance is not feasible.

We conclude that $MCLCP$ with three or more carriers and non-identical cost structures is NP-Hard. $\square$

APPENDIX B

PROCEDURE FOR FINDING A NESTED MOAT PACKING

Suppose that we solve the dual MP and obtain a maximum moat packing vector w for the TSP that does not have a nested structure. This means that there are at least two moats that intersect. Let the width of the moat surrounding a subset S of customers be denoted by w_S in this discussion.

Let S_1, S_2, and S_3 be three disjoint partitions of the set $N \cup \{0\}$ and let $w_{S_1 \cup S_3} > 0$ and $w_{S_2 \cup S_3} > 0$ be the two moats that intersect and let $\epsilon = \min\{w_{S_1 \cup S_3}, w_{S_2 \cup S_3}\}$. The procedure updates the width of the moats and obtains another moat packing vector $\bar{w}$ such that

$$\bar{w}_{S_1} = w_{S_1} + \epsilon, \quad \bar{w}_{S_2} = w_{S_2} + \epsilon,$$

$$\bar{w}_{S_1 \cup S_3} = w_{S_1 \cup S_3} - \epsilon, \quad \bar{w}_{S_2 \cup S_3} = w_{S_2 \cup S_3} - \epsilon, \quad \bar{w} = w \ \ \forall \ others.$$

We claim that $\bar{w}$ is a feasible optimal solution for linear program MP. The proof of optimality is straight forward. The vector w is a maximum moat packing vector and we increase the width of two moats (w_{S_1}, w_{S_2}) and decrease the width of two moats ($w_{S_1 \cup S_3}$, $w_{S_2 \cup S_3}$), hence the vector $\bar{w}$ is also a maximum moat packing vector. The constraints of linear program MP is not violated by $\bar{w}$, since both moats (w_{S_1}, w_{S_2}) are included in two sets of constraints and in both constraint sets one of the moats ($w_{S_1 \cup S_3}$, $w_{S_2 \cup S_3}$) is also included. Hence, the left hand sides remain unchanged as can be seen below.

$$w_{S_1} + \cdots + w_{S_2 \cup S_3} + \cdots \leq g_{ij} \quad \forall i \in S_1, j \in S_3 \tag{88}$$

$$w_{S_1} + \cdots + w_{S_1 \cup S_3} + \cdots \leq g_{ij} \quad \forall i \in S_1, j \in (N \cup \{0\}) \backslash (S_1 \cup S_3) \tag{89}$$

$$w_{S_2} + \cdots + w_{S_1 \cup S_3} + \cdots \leq g_{ij} \quad \forall i \in S_2, j \in S_3 \tag{90}$$

$$w_{S_2} + \cdots + w_{S_2 \cup S_3} + \cdots \leq g_{ij} \quad \forall i \in S_1, j \in (N \cup \{0\}) \backslash (S_2 \cup S_3). \tag{91}$$

Therefore, we conclude that $\bar{w}$ is a feasible optimal solution for linear program MP. Note that the moat packing $\bar{w}$ has one few intersection of moats compared to w. By repeating the procedure as needed, we finally obtain a moat packing that has a nested structure. Note that the procedure may require to handle an exponential number of variables in every step. Therefore, the overall computational time of the procedure is not polynomial.

REFERENCES

[1] Ahuja, R. K., Magnanti, T. L., and Orlin, J. B., *Network Flows : Theory, Algorithms, and Applications.* Englewood Cliffs, N.J.: Prentice Hall, 1993.

[2] An, N., Elmaghraby, W., and Keskinocak, P., "Bidding strategies and their impact on revenues in combinatorial auctions," *J Revenue Pricing Manag*, vol. 3, no. 4, pp. 337–357, 2005.

[3] ATA, "Fuel talking points," tech. rep., American Trucking Associations. http://www.trucking.org, Date Accessed: September 20, 2007.

[4] ATA, "Trucking activity report," tech. rep., American Trucking Associations. http://www.trucking.org, Date Accessed: September 4, 2005.

[5] Bertazzi, L., Savelsbergh, M. W. P., and Speranza, M. G., "Inventory routing." 2008.

[6] Bertsimas, D. and Tsitsiklis, J. N., *Introduction to Linear Optimization.* Athena Scientic, 1997.

[7] Borm, P., Hamers, H., and Hendrickx, R., "Operations research games: A survey," *TOP*, vol. 9, no. 2, pp. 139–199, 2001.

[8] Campbell, A. M., Clarke, L., Kleywegt, A., and Savelsbergh, M. W. P., "The inventory routing problem," in *Fleet Management and Logistics* (Crainic, T. and Laporte, G., eds.), Boston: Kluwer Academic Publishers, 1998.

[9] Caplice, C. and Sheffi, Y., "Optimization-based procurement for transportation services," *Journal of Business Logistics*, vol. 24, no. 2, pp. 109–128, 2003.

[10] Cordeau, J.-F., Laporte, G., Savelsbergh, M. W. P., and Vigo, D., "Short-haul routing," in *Handbooks in Operations Research and Management Science: Transportation* (Laporte, G. and Barnhart, C., eds.), 2007.

[11] De Vries, S. and Vohra, R. V., "Combinatorial auctions: A survey," *Informs Journal on Computing*, vol. 15, no. 3, pp. 284–309, 2003.

[12] Derks, J. and Kuipers, J., "On the core of routing games," *International Journal of Game Theory*, vol. 26, no. 2, pp. 193–205, 1997.

[13] Elmaghraby, W. and Keskinocak, P., "Combinatorial auctions in procurement," in *The Practice of Supply Chain Management: Where Theory and Application Converge*, pp. 245–258, 2004.

[14] Engelbrecht-Wiggans, R. and Granot, D., "On market prices in linear production games," *Mathematical Programming*, vol. 32, no. 3, pp. 366–370, 1985.

[15] Engevall, S., Gothe-Lundgren, M., and Varbrand, P., "The traveling salesman game: An application of cost allocation in a gas and oil company," *Annals of Operations Research*, vol. 82, no. 0, pp. 203–218, 1998.

[16] Engevall, S., Gothe-Lundgren, M., and Varbrand, P., "The heterogeneous vehicle-routing game," *Transportation Science*, vol. 38, no. 1, pp. 71–85, 2004.

[17] Ergun, O., Kuyzu, G., and Savelsbergh, M., "Reducing truckload transportation costs through collaboration," *Transportation Science*, vol. 41, no. 2, pp. 206–221, 2007.

[18] Ergun, O., Kuyzu, G., and Savelsbergh, M., "The shipper collaboration problem," *Computers and Operations Research*, vol. 34, no. 6, pp. 1551–1560, 2007.

[19] Faigle, U., Fekete, S. P., Hochstattler, W., and Kern, W., "On approximately fair cost allocation in euclidean tsp games," *Or Spektrum*, vol. 20, no. 1, pp. 29–37, 1998.

[20] Goemans, M. X. and Skutella, M., "Cooperative facility location games," *Journal of Algorithms*, vol. 50, no. 2, pp. 194–214, 2004.

[21] Gothe-Lundgren, M., Jornsten, K., and Varbrand, P., "On the nucleolus of the basic vehicle routing game," *Mathematical Programming*, vol. 72, no. 1, pp. 83–100, 1996.

[22] Granot, D., Hamers, H., and Tijs, S., "On some balanced, totally balanced and submodular delivery games," *Mathematical Programming*, vol. 86, no. 2, pp. 355–366, 1999.

[23] Immorlica, N., Mahdian, M., and Mirrokni, V. S., "Limitations of cross-monotonic cost sharing schemes," in *Proceedings 16th ACM Symposium on Discrete Algorithms*, pp. 602–611, 2005.

[24] Jain, K. and Vazirani, V. V., "Applications of approximation algorithms to cooperative games," in *Proceedings 33rd ACM STOC*, pp. 364–372, 2001.

[25] Kalai, E. and Zemel, E., "Generalized network problems yielding totally balanced games," *Operations Research*, vol. 30, no. 5, pp. 998–1008, 1982.

[26] Kuipers, J., "A note on the 5-person traveling salesman game," *Mathematical Methods of Operations Research*, vol. 38, no. 2, pp. 131–139, 1993.

[27] LEDYARD, J. O., OLSON, M., PORTER, D., SWANSON, J. A., and TORMA, D. P., "The first use of a combined-value auction for transportation services," *Interfaces*, vol. 32, no. 5, pp. 4–12, 2002.

[28] MOORE, E. W., WARMKE, J. M., and GORBAN, L. R., "The indispensable role of management science in centralizing freight operations at reynolds-metals-company," *Interfaces*, vol. 21, no. 1, pp. 107–129, 1991.

[29] MOULIN, H., *Cooperative Microeconomics : A Game-Theoretic Introduction.* Princeton, N.J.: Princeton University Press, 1995.

[30] OSBORNE, M. J. and RUBINSTEIN, A., *A Course in Game Theory.* The MIT Press, 1994.

[31] OWEN, G., "On the core of linear production games," *Mathematical Programming*, vol. 9, no. 3, pp. 358–370, 1975.

[32] PAL, M. and TARDOS, E., "Group strategyproof mechanisms via primal-dual algorithms," in *Proceedings of the Annual ACM Symposium on the Theory of Computing*, 2003.

[33] POTTERS, J. A. M., CURIEL, I. J., and TIJS, S. H., "Traveling salesman games," *Mathematical Programming*, vol. 53, no. 1, pp. 199–211, 1992.

[34] RUCHIRA, S. D., "Using computer algebra to find nash equilibria," in *Proceedings of the 2003 international symposium on Symbolic and algebraic computation*, (Philadelphia, PA, USA), ACM, 2003.

[35] SAMET, D. and ZEMEL, E., "On the core and dual set of linear-programming games," *Mathematics of Operations Research*, vol. 9, no. 2, pp. 309–316, 1984.

[36] SANCHEZ-SORIANO, J., LOPEZ, M. A., and GARCIA-JURADO, I., "On the core of transportation games," *Mathematical Social Sciences*, vol. 41, no. 2, pp. 215–225, 2001.

[37] SCHMEIDLER, D., "Nucleolus of a characteristic function game," *Siam Journal on Applied Mathematics*, vol. 17, no. 6, pp. 1163–1170, 1969.

[38] SHAPLEY, L. S., "A value for n-person games," *Contributions to the Theory of Games, II, Annals of Mathematical Studies*, vol. 28, pp. 307–317, 1953.

[39] SHEFFI, Y., "Combinatorial auctions in the procurement of transportation services," *Interfaces*, vol. 34, no. 4, pp. 245–252, 2004.

[40] SONG, J.-H. and SAVELSBERGH, M. W. P., "Performance measurement for inventory routing," *Transportation Science*, vol. 41, no. 1, pp. 44–54, 2007.

[41] SONG, J. J. and REGAN, A., "Combinatorial auctions for transportation service procurement - the carrier perspective," in *Freight Policy, Economics, and Logistics; Truck Transportation*, Transportation Research Record, pp. 40–46, 2003.

[42] Strozniak, P., "Collaborative logistics: Overcoming its challenges can lower transportation and inventory costs and reduce stockouts," *Frontline Solutions*. http://www.frontlinetoday.com, Date Accessed: June 6, 2005.

[43] Tamir, A., "On the core of a traveling salesman cost allocation game," *Operations Research Letters*, vol. 8, no. 1, pp. 31–34, 1989.

[44] Tamir, A., "On the core of network synthesis games," *Mathematical Programming*, vol. 50, no. 1, pp. 123–135, 1991.

[45] Tazari, S., "Crossmonotonic cost sharing schemes for combinatorial optimization games: A survey." University of British Columbia, 2005.

[46] TC, "Transportation in canada," tech. rep., Transport Canada. http://www.tc.gc.ca, Date Accessed: September 7, 2007.

[47] Yahoo, "Industry center - trucking industry overview," tech. rep., Yahoo Finance. http://biz.yahoo.com/ic/774.html, Date Accessed: September 20, 2007.

[48] Young, H. P., *Cost Allocation: Methods, Principles, Applications (Ed.)*. Amsterdam, The Netherlands: North-Holland, 1985.