

Technical University of Denmark



Service Network Design and Management in Linear Container Shipping Applications

Andersen, Martin Wahl; Madsen, Oli B.G. ; Stidsen, Thomas Jacob Riis

Publication date:
2010

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Andersen, M. W., Madsen, O. B. G., & Stidsen, T. R. (2010). Service Network Design and Management in Linear Container Shipping Applications. Kgs. Lyngby, Denmark: Technical University of Denmark (DTU).

DTU Library

Technical Information Center of Denmark

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Service Network Design and Management in Liner Container Shipping Applications

Martin W. Andersen

February 2010

Service Network Design and Management in Liner Container Shipping Applications

Martin W. Andersen

Kongens Lyngby 2010

PHD-2010-06
Department of Transport
Technical University of Denmark

Copyright: Copying permitted if source is stated.

Published by: DTU Transport
Bygningstorvet 116 Vest
DK-2800 Kongens Lyngby
Denmark

Order at: <http://www.transport.dtu.dk>
Tel +45 45256500, Fax +45 45936412
transport@transport.dtu.dk

ISBN: 978-87-7327-210-7 (Electronic version)

ISBN: 978-87-7327-211-4 (Printed version)

Abstract

The central problem treated in this thesis is that of designing and maintaining a service network in a liner container shipping context. Based on a unified description of the individual planning processes involved in the definition of the liner container service network design (SND) problem, a series of integrated models for the SND problem are developed in an iterative fashion. The first models that are proposed represent simple abstractions of the SND problem but gradually, additional dimensions are added to obtain rich models capable of capturing many of the requirements imposed on the liner container SND.

Two concrete problems, the *liner container feeder service network design* problem and the *network transition* problem are presented and analyzed in further detail with the purpose of providing tools to support the planning processes related to the design of the service network as well as the realization of a new design.

The first problem, the *liner container feeder service network design* problem, addresses the tactical planning of the service network. The problem is inspired by a real-world case and problem specific structural properties are exploited to develop a new decomposition strategy. Essentially, the problem is decomposed into two types of sub-problems; a route generation problem and a route packing problem. The key to the success of this strategy is the introduction of the concept of a route pool which is iteratively augmented with route candidates using a dual based heuristic. Furthermore, dual estimation is used to select attractive routes from the route pool. Selected routes are managed in a master problem which is dynamically expanded as new routes and route packings are generated. Through the proposed decomposition strategy, a very rich representation

of the liner service network design problem can be achieved modeling complex aspects such as service level dependent demand.

Computational experiments are conducted on a series of instances based on a real-world case. Results show that the proposed solution approach is capable of solving realistically sized problems. Furthermore, solutions produced are of a high quality achieving consistently high capacity utilization and the data owner has shown interest in performing additional analysis of the produced results.

The *network transition* problem is a new problem not previously treated in the literature. The problem seeks to reduce the barrier of adopting advanced techniques to perform service network design by addressing the problem of migrating a fleet of vessels from one service network design to another. The problem shares features with the well known pickup and delivery problem but extends this in several ways.

A parallel cooperative adaptive large neighborhood search (ALNS) based heuristic is proposed to solve the network transition problem. The heuristic is based on a ruin and recreate principle to perform neighborhood moves using both well known as well as problem specific neighborhood operators. An adaptive mechanism selects neighborhood operators based on their previous performance. The ALNS framework is quite general and is particularly well suited for highly constrained problems where traditional local search methods based on small neighborhood moves have difficulty moving between regions of the solution space.

The ALNS is analyzed through a series of computational experiments and is shown to be quite robust against different settings for the algorithm parameters. In terms of scalability, the ALNS is shown to be capable of solving problem instances of up to 400 commodities (800 requests the terminology used in the pickup and delivery literature) within a time frame reasonable to the planning problem it supports. To evaluate the behavior of the ALNS in a realistic scenario, an instance adapted from a real-world case using historical data is created. Results show that the ALNS can offer a savings potential of more than \$100,000 compared to the currently operated schedule.

Resumé

Design og vedligeholdelse af et service netværk inden for den liner baserede container shipping industri er den central problemstilling der behandles i denne afhandling. Baseret på en beskrivelse af de individuelle planlægningsprocesser der er forbundet med bestemmelsen af et liner container service netværk udvikles iterativt en serie integrerede modeller til beskrivelse af service netværksdesign problemet. De første modeller i denne serie er simple abstraktioner af service netværksdesign problemet, men yderligere dimensioner bliver gradvist tilføjet for til sidst at opnå detaljerede modeller i stand til at afspejle mange af de krav der stilles til designet af et liner container service netværk.

To konkrete problemer, liner container feeder service netværks design problemet og netværkstransitionsproblemet, fremhæves og bliver analyseret i flere detaljer. Formålet med disse to problemer er at bidrage med værktøjer til at støtte planlægningsprocesser relateret til design af service netværk inden for liner container shipping samt realiseringen og implementeringen af disse nye design.

Det første problem, *liner container feeder service netværksdesign problemet*, behandler den taktiske planlægning af et service netværk. Problemet er inspireret af et reel problemstilling fra industrien og strukturelle egenskaber bliver udnyttet i udviklingen af en løsningsmetode baseret på en ny strategi til at dekomponere problemet. Grundlæggende bliver problemet opdelt i et overordnet problem samt to typer af underproblemer, et rutegenereringsproblem og et rutestuvningsproblem. Nøglen til at opnå success ved denne dekompositionsstrategi ligger i introduktionen af en rutepulje som iterativt udvides med nye rutekandidater baseret ved hjælp af en dual-baseret heuristik. Endvidere anvendes

dual estimation til at udvælge attraktive ruter fra rutepuljen. Disse udvalgte ruter styres via det overordnede problem og dette problem udvides dynamisk efterhånden som nye ruter og rutestuvninger genereres. Ved at anvende den foreslåede dekompositionsstrategi opnås en særdeles detaljeret repræsentation af liner service netværksdesign problemet der giver mulighed for at modellere komplekse aspekter som serviceafhængige fragtvoluminer.

Eksperimenter udført for en række datasæt baserede på et case fra industrien viser at den foreslåede løsningsmetode er i stand til at håndtere problemer af realistisk størrelse. Envidere viser eksperimenter at løsningsmetoden er i stand til at producere service netværk af så høj en kvalitet og med høj kapacitetsudnyttelse at problemejeren har vist interesse i at foretage yderligere analyser af de foreslåede løsninger.

Det andet problem kaldet *netværkstransitionsproblemet* er et nyt problem som ikke tidligere er behandlet i litteraturen. Problemet behandler udfordringen forbundet med at migrere en flåde af skibe fra et service netværk til et nyt. En løsning af denne problemstilling har til formål at nedbryde en af de barrierer der pt er for indførelsen af avancerede teknikker til understøttelse af planlægningsprocesser relateret til design af et liner service netværk.

En heuristik baseret på parallel adaptiv nabolagssøgning er implementeret til løsning af netværkstransitionsproblemet. Heuristikken baserer sig på en nedbryd-og-genopbyg princip og anvender både velkendte samt problemspecifikke nabolagsalgoritmer i søgningenstrategien. En adaptiv mekanisme udvælger nabolagsalgoritmer baseret på deres hidtidige success. Den foreslåede løsningsmetode er generel i sin implementering og er særligt velegnet til meget restriktive problemstillinger hvor traditionelle lokalsøgningsalgoritmer kan have vanskeligt ved at bevæge sig mellem forskellige regioner i løsningsrummet.

Gennem en række eksperimenter analyseres den parallelle adaptive nabolagssøgningsalgoritme og det vises at algoritmen er robust over for forskellige indstillinger af input parametre. Det vises også at algoritmen skalerer tilfredsstillende og er i stand til at løse problemer med op til 400 fragtemner inden for en tidsgrænse der vurderes som værende realistisk i forhold til den planlægningsprocess som den er designet til at understøtte. Endelig evalueres algoritmen i et mere realistisk scenario baseret på data fra et case fra industrien og resultater viser at potentielle besparelser på mere end \$100,000 kan opnås sammenlignet med den oprindelige sejlplan i de underliggende data.

Preface

Attempting to write a PhD thesis addressing problems and their optimal solutions represents a conflict at its core. On one hand, optimization deals with a pursuit of perfection; a strive to obtain the best possible solutions, referred to as “optimal” solutions, to some given problem relative to some pre-defined measure of quality. These problems typically involve one or several resource constraints subject to which the optimal solution must be established and the measure of quality is typically well defined (at least in the realm of the problem abstractions considered). On the other hand, achieving perfection in documenting the ideas behind obtaining such optimal solutions is in itself a resource constrained problem in which the structure allows for infinite permutations and where the measure of quality is multi-dimensional and much more difficult to evaluate. Thus the task of writing this thesis necessarily became a problem of deciding when a satisfactory structure, depth, and quality had been achieved subject to what initially may have seemed a fairly loosely (time) constrained problem which quickly became highly resource constrained. This thesis represents my ultimate decisions regarding structure, depth, and content in answering some of many questions and challenges that lie ahead to bring structured and informed problem solving to the Liner Container Shipping Industry.

This thesis has been prepared at the Department of Transport, Technical University of Denmark in partial fulfillment of the requirements for the degree Doctor of Philosophy PhD in engineering science.

The problems discussed in this theses are based on interviews and collaboration with a major container feeder service provider and enriched by the knowledge obtained through several years of experience working

closely with actors in the liner container shipping industry.

Kgs. Lyngby, Denmark, January 2010
Martin W. Andersen

Acknowledgments

A PhD study is an endeavor that should not be undertaken lightly. It requires great dedication and sometimes also personal sacrifices to successfully reach the goal. What one might sometimes fail to realize (or simply temporarily forget) is that research is an iterative process in which progress is not necessarily achieved at each step and often, the best sequence of steps is not fully known at the outset. Being inexperienced with this process requires a very high level of perseverance in the face of the inevitable setbacks that one is faced with during the course of a PhD study. Completing such a tremendous work is a lonely journey but a large factor in ultimately ensuring success is the people that we have around us.

First of all, I would like to thank my supervisor Professor Oli B.G. Madsen for his unwavering confidence in my abilities to eventually complete this thesis, continued belief in my skills in self-management, and for his guidance and support during the creation of this work.

I would also like to thank Professor Cynthia Barnhart for her support and knowledgeable input during my stay at the ORC at MIT. I am very grateful for the many fruitful discussions we have had and her suggestions have been a significant factor in determining the direction and focus of my thesis.

Additional thanks go to Thomas Stidsen for inspiring and fruitful discussions about matters related to both the problems discussed in this thesis as well as refreshing sidetracks into unrelated areas, always with an alternate perspective on things.

Furthermore, I would also like to thank the skilled people at company

providing data who have been a great source of in-depth and updated knowledge about the liner container industry. Without their willingness to supply data and problem insights, many of the problems discussed in this thesis would not have been so rich.

And to you Christina; thank you for showing endless patience and understanding during periods of high stress levels and for taking care of our lovely son; his laughs and smiles go a long way toward brightening up even the most gray days.

Contents

1	Introduction	1
1.1	Global Freight Transportation	3
1.2	Service Networks and Supply Chains	6
1.3	Contributions	8
1.4	Thesis Overview	9
2	Liner Container Shipping	11
2.1	Current and Future Trends	13
2.2	Planning Levels, Decisions and Objectives	16
2.3	Service Network Components	23
2.4	Differences in Short Sea and Deep Sea Operating Environments	31
2.5	Network Transition and Recovery	32
3	Literature Review	37
3.1	Simulation Based Approaches	40
3.2	Decomposition/Relaxation Based Approaches	41
3.3	Column Generation Approaches	42
3.4	Related Literature	43

3.5	Summary	45
4	Modeling Liner-Based Service Networks	47
4.1	Multicommodity Capacitated Network Design	50
4.2	Classical Network Design Extended	54
4.3	Decomposing using Specific Problem Properties	61
4.4	Models with a Set Partitioning Structure	66
4.5	Summary	69
5	Paper 1: Feeder Service Network Design	71
5.1	Introduction	72
5.2	Liner Container Feeder Planning	75
5.3	Mathematical Model	78
5.4	Solution Methodology	85
5.5	Computational Experiments	90
5.6	Concluding Remarks	97
6	Paper 2: Network Transition	101
6.1	Introduction	102
6.2	The Network Transition Problem	105
6.3	Solution Approach	109
6.4	Computational Experiments	118
6.5	Conclusions and Further Work	129
7	Comments on Papers	135
7.1	Paper 1: Liner Feeder Service Network Design	135
7.2	Paper 2: Network Transition in a Liner Application	136
8	Conclusion	143
	Bibliography	147

CONTENTS	xi
A Balanced Network Design	157
A.1 Relaxation of the BMCND	158
A.2 The Lagrangian Dual Problem	161
A.3 Heuristic and Branch-and-Bound	164
A.4 Preliminary Computational Experiments	164
B Data Generator	167
B.1 Geography	168
B.2 Commodities	169
B.3 Schedules	171
B.4 Parameter Overview	174
C Glossary and Abbreviations	177

List of Figures

1.1	Overview of major freight transport modes	4
2.1	Overview of cargo handling before/after containerization .	12
2.2	Planning levels and decisions.	17
2.3	Major topology types in liner networks	28
2.4	Full and empty container share, transshipments.	29
4.1	Multi-layer asset balanced network	58
5.1	Time-space representation of the service network	80
5.2	Example of two vessels operating between four hubs.	80
5.3	Algorithmic flow.	86
6.1	Network transition example	107
6.2	Overview of subsequence removal algorithm	115
6.3	Objective development, multi-thread comparison	124
6.4	Example of the development of the objective fuction value.	128
B.1	Port layout with a hub region	169
B.2	Port distribution in a random layout	169

B.3	Distribution of time windows (example)	170
B.4	Example of a generated schedule	172

List of Tables

2.1	Feeder vs. deep sea operating environment	32
3.1	Overview of literature	39
5.1	Real-world scenario characteristics	91
5.2	Results of baseline experiments	92
5.3	Results from route enumeration and dual selection	94
5.4	Results from applying dominance removal	95
5.5	Results from heuristic route pool augmentation	96
6.1	Overview of tuning instances.	121
6.2	Effect of neighborhood size	121
6.3	Performance of insertion heuristics	123
6.4	Speed-up factor using multiple threads	124
6.5	Results for the Small-class test instances.	126
6.6	Results for the Medium-class test instances.	126
6.7	Results for the Large-class test instances.	127

CHAPTER 1

Introduction

“...it often now costs more to ship a container by road 100 miles from a port to its final destination than it does to move the container by sea from China to Europe.” Wright (2006)

The above quote encompasses one of the key factors responsible for the incredible success and interesting development that the liner container shipping industry has seen and facilitated during its relatively short history. It is a testament to the magnificent and deceptively simple technology that the container is and to the skillful execution of the movement of these units of cargo. Quite simply, containerized freight transportation has revolutionized the modern supply chain end to end. From the sourcing of materials and parts to the distribution and expansion of products on small and large markets around the world.

With an estimated annual growth rate of 10% during the past two decades, containerized trade has expanded from a share of 5.1% of the world's total dry cargo trade in 1980 to 25.4% in 2008, UNCTAD (2009). This explosive growth is driven primarily by the globalization of trade and outsourcing of production to traditional low wage countries, mainly in Asia with a concentration in the Far East. Developments in political (e.g., deregulation) and economical (e.g., open markets) environments as well as the increases in volume and speed of information exchange (business-to-business) continue to support these trends. Although only

representing a relatively small fraction of the total world trade measured in tonnes ($\sim 16\%$), the value of containerized trade grew to \$4 trillion in 2008 representing more than 70% of the value of seaborne trade, WTO (2008). Based on these statistics it is clear that containerized trade plays a significant role in the global trade and economy.

With such a large and rapidly growing market, new entrants are a constant threat to established carriers and their presence intensify the already fierce competition on the market. Achieving economy of scale is crucial to the success of a carrier as it is an important factor in being able to compete on service and price, by far the two dominant factors in attracting freight. Although there are many smaller players in the liner container freight forwarding market, recent years have seen a series of mergers and acquisitions such that a few handfuls of very large players now carry more than 50% of the total volume.

As a result of constant competition, growth, consolidation, and customer demand for specialized freight forwarding services, liner carriers, and specifically liner container carriers, operate large and complex consolidation based service networks. The term service network generally refers to a network consisting of a series of terminals at which (containerized) freight can be handled and a set of vessels performing freight forwarding services between these terminals. To meet changing market conditions (e.g., demand and actions of competitors), carriers, regardless of size, constantly need to revise their service networks. This is no simple task but in an industry suffering from constantly decreasing profit margins, designing and operating the service network as efficiently as at all possible taking into consideration multidimensional constraints and requirements is crucial to the future success and survival of a carrier.

This thesis addresses the problem of designing and managing a liner based service network in the specific context of liner container shipping. This problem will be referred to as the liner service network design problem or simply the service network design (SND) problem. As it will become clear in the following chapters, SND is a complex and central planning process of liner carriers, but is still mostly executed through manual procedures by experienced planners. This thesis approaches the SND problem from the perspective of operations research, showing possible avenues in the process toward developing models, algorithms and systems to help support the SND planning process. An important aspect of this work is to develop an integrated business understanding to fully evaluate the merits and limitations of developed models. Central to the developments in this thesis is a recognition that to remain competitive in a global and changing market, carriers must possess the ability to adapt

and innovate to stay ahead. The speed at which changes happen is increasing as are demands on the service networks. Entirely new thinking and approaches to the SND is thus required to keep ahead in such a market. In this perspective, to leverage the potential benefits of the models proposed in this work, both technological and managerial innovation is required on the part of carriers choosing to adopt new tools in their SND planning processes. This work addresses a part of the technological aspect.

1.1 Global Freight Transportation

Modern global freight transportation is carried out over complex networks with a large number of interfaces and routing alternatives. In contrast to local and regional freight transportation which is typically limited to a single routing mode (e.g., trucks), global international/intercontinental freight is often carried using a combination of different modes which may include truck, rail, air, and ship. This combination of several modes in the door to door transport of individual commodities leads to the need for efficient means of intermodal transfer of freight.

Consolidation networks in which freight can easily be transferred between different modes provide opportunities to take advantage of economy of scale in the selection of the mode of transport for the individual legs of an intermodal journey. For example, the intercontinental door to door journey of a containerized commodity will typically begin by being picked up by a truck from the shipper (sender) and transferred to a port terminal. Here it is consolidated with other commodities and transferred to a regional feeder ship transporting these commodities to a regional hub where it is again consolidated and transshipped to a deep sea ship executing the main leg of the journey. At a hub in the destination region the commodity is discharged, and then with decreasing levels of consolidation first transferred onto a feeder and then moved via truck to the final destination at the consignee (receiver). In this example, and in fact in much of the global transport of produced or semi-produced commodities, sea-based transport of containerized goods play a significant role in realizing efficient consolidation networks.

Apart from geographical and infrastructural constraints limiting the use of a single mode of transport in long-haul freight transports, there are also cost and time incentives for the use of multiple modes as briefly suggested in the above example. While trucks are very flexible, relatively fast, and can reach most locations they have only a limited capacity and

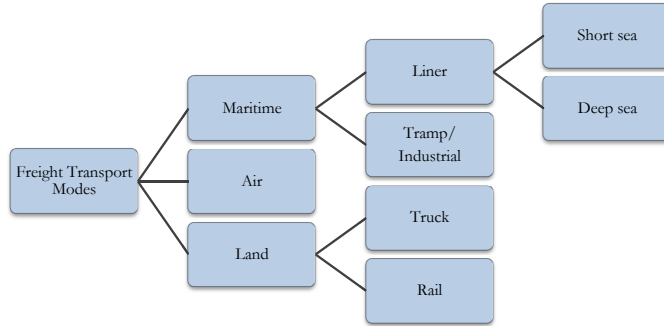


Figure 1.1: Overview of major freight transport modes and selected sub-modes.

are relatively costly. Rail on the other hand, can carry larger volumes of freight but suffer from rigid, limited infrastructure, slow service, and is limited to operate on the same continent (or even in the same country due to equipment incompatibilities) just as trucks. For high speed intercontinental transport, air is the mode of choice but its limited capacity and high cost means it is mostly used to carry low volume time sensitive freight (e.g., express package shipments). Finally, ships offer high volume, low speed intercontinental freight transport service at a very low cost (relative to other modes). Christiansen et al. (2004) provides an additional comparison of the operational characteristics for the above four modes of transportation. An overview of major transport modes together with relevant sub-types is provided in Figure 1.1 and the maritime modes will be discussed further in the following section.

1.1.1 Sea-based Freight Transportation

With respect to sea-based freight transportation, there is a typical distinction between three primary modes of operation; industrial, tramp and liner shipping. Modes define how the freight forwarding services are published and executed but do not determine the type of cargo that is carried nor the particular vessel design that is operated. However, characteristics of different cargo types naturally limit the modes under which these can be profitably moved. *Industrial* and *tramp* shipping are very similar as both modes operate schedules that are dynamic and based on the near-term availability of freight. Ships simply go where the freight is. Such an operation essentially requires freight of a large volume relative to the vessels capacity to be profitable. For this reason, tramp and

industrial shipping deals mainly with liquid (e.g., oil and liquefied gas) and bulk (e.g., minerals and grains) products. Industrial shipping differs from tramp shipping in terms of fleet ownership as the shipper is also the fleet owner. This mode of operation is on the decline and now mostly seen with large oil/gas companies. *Liner* shipping is based on a fixed and published schedule in which vessels operate cyclic routes (referred to as rotations). A liner carrier offers a series of services (or products) which together with the published schedule provide complete transport itineraries for shippers. Liner carriers do not require any specific relationship with the shippers and are as such referred to as common carriers. Lawrence (1972) contains a more extensive definition and comparison of the three modes of operation from both a technical and operational perspective.

Within the class of liner shipping, there is often a distinction between *short sea* and *deep sea* operations. Although a single carrier may operate both types of service, it is increasingly common that carriers focus their core business and offer primarily one type of service. In this setup, short sea carriers will service both intra-region freight as well as provide feeder service for the deep sea carriers who then handle the main haul, typically over a longer distance. A series of major ports referred to as hubs will provide the interface between the service networks of short sea and deep sea carriers through the transshipment of freight. These hubs can also be thought of as consolidation points with the highest level of consolidation generally achieved on the deep sea main haul leg, as in the example provided in the previous section.

It is worth noting that although cost of shipping containerized freight using liner services has fallen dramatically since its inception, it remains one of the most expensive forms of maritime freight transport measured per weight or volume. Handling containers incurs significant costs to the transport which also means that freight carried by liner container services is typically high value, low volume commodities such as parts or finished consumer products. This is to be contrasted with e.g., bulk products such as coal or food grains that have a relatively low value per volume unit. Despite the dominance of high value goods in liner container freight, the industry still suffers under constantly diminishing freight rates resulting in increasingly lower margins. Thus, goals such as market differentiation, cost reductions, and efficiency increase will be a recurring theme of this thesis. The means to achieve these goals are high asset utilization and service network optimization which will be the subject of the remainder of this work. More specifically, focus will be on the design and management of the service network for liner container carriers.

Emphasizing the broad importance of sea/waterway based transport is the role it plays in developing countries where it is often the only really viable mode of transportation due to poor availability of suitable inland transport infrastructure such as road or rail networks. As a result, trade becomes concentrated near ports and landlocked countries face a high barrier for international/intercontinental trade in the form of added cost of transport and lower reliability of the logistics chain. Interestingly, even in highly developed regions like the European Union, congestion on inland road and rail infrastructure has led to the increased focus on the transfer of freight to short sea and inland waterway shipping modes (EC, 2008). Hindering efforts toward such a shift is the current efficiency and capacity problems at ports and to some extent also the relatively poor *interfaces* between sea and land based transport modes. In this context, Paixão and Marlow (2002) provides an analysis of the strengths and weaknesses of short sea shipping and Notteboom and Rodrigue (2008) provides a further discussion of issues relating to intermodal integration with specific emphasis on the port-liner network interface. Together with improvements to ports and modal interfaces, continued work toward ensuring a common regulatory system is also necessary for the future success of short sea as well as deep sea shipping.

1.2 Service Networks and Supply Chains

To understand and set into a larger context the problems that are addressed in this thesis it is useful to distinguish between two major network types; transmission networks and transport or service networks. In *transmission networks*, the network itself is represented by physical installations on links between a series of nodes (locations) over which transmission occurs. Examples of transmission networks include electrical power networks, water and sewage grids, and telecommunications networks. Specific for telecommunications networks is that links are generally bi-directional.

Service networks are characterized by a set of assets operating a series of services over a network which consists of a number of terminals connected either by conceptual (sea/air) or physical (rail/road) links. Services are typically based on published schedules and offered by common carriers, i.e., without any specific requirements to the relationship between shipper and carrier. Assets are moving between the nodes of the network thus creating the services which is in contrast to the transmission networks where assets are stationary. The main requirement in service networks is that assets balance during their services to ensure that services can be

continually operated. Furthermore, in service networks, the goods that flow on the network will usually induce flows in opposite directions as a result of reverse logistics. This is true in trucking where individual vehicles return to a depot after ending their route and in liner container shipping where empty containers need to be repositioned to the locations that demand these.

Service networks are subject to the same supply and demand mechanisms governing in many other open markets. Producers of goods require freight forwarding services to satisfy their distribution and sourcing requirements and thus represent the demand side. On the supply side, *carriers* offer transport services that may be either customized (as is the case for e.g., tramp shipping) or consolidation based with predefined schedules (as in liner shipping). In order for a carrier to successfully satisfy demand for freight forwarding services, it must continually redesign and adjust its service network in response to the market. The frequency and scope of such redesigns is determined by the specific type of service network and determines the methods used to aid the design.

Due to the high complexity and system wide correlation, service network planning is a particularly obvious candidate for the use of advanced planning methods such as operations research (OR). One classical example of a successful adoption of OR in network planning is the case of the airline industry. In the mid 1970's, airline companies were facing deregulation which combined with increasing fleet sizes resulted in highly complex planning processes involving continually larger human resources. Also, due to the size of the network, planning processes were limited to considering only smaller parts of the whole service network leading to inefficient routing and scheduling plans. Organizational resistance initially hindered the full implementation of automated planning processes, but strong managerial drive eventually ensured success to such a level that today, planning in the airline industry is considered impossible without the use of these planning systems.

Similar trends are emerging in the liner shipping industry and recent years have seen an increasing interest in and awareness of the need for improvements in the network planning processes. The recognition of this need is driven by some of the same problems as those originally faced by e.g., the airline industry; increasing fleet sizes and network complexity. In terms of network complexity, the challenges include a continued increase in demand for reliable, secure and fast services in a market that at the same time is facing constantly decreasing profit margins. Adding to the pressure is the growing international focus on green projects and the reduction of CO₂ emissions. Finally, customers (shippers) face the same

kind of global price pressures and competition forcing them to look at optimizing their own processes. Here, the supply chain plays a big role as it affects many processes of e.g., manufacturing companies and thus is a prime candidate for optimization. Concepts such as just-in-time and lean no longer provide a competitive advantage but are rather a requirement to operate. This focus on supply chains and their integration naturally affects how carriers are expected to operate as well as the type and quality of services they offer. To this end, services are not only thought of as freight forwarding services but now also include information sharing both before, during and after business transactions.

1.3 Contributions

The objective of this thesis is to provide an introduction to and analysis of the major challenges that liner container shipping service providers face when deciding to adopt more advanced methods based on the principles of operations research in their service network related planning processes. Contributions fall in two parts. The first part addresses the liner service network design problem by highlighting and describing the primary features and constraints that apply in a selected subset of planning processes. Based on the highlighted features, a series of models are compared and evaluated from the perspectives of algorithmic challenges and limitations with respect to modeling the service network design problem. Modeling approaches evaluate and integrate relevant ideas and methods from related problem areas.

In the second part, two specific problems; the *container feeder service network design* problem and the *network transition* problem, are addressed in greater detail and analyzed through extensive computational experiments using both synthetic and real-world data. Together, the two proposed solution approaches support and link planning processes from the initial network design to the final realization in an operational network. It is worth emphasizing that the network transition problem has not previously been addressed in the literature as far as the author has been able to determine. Also, the decomposition ideas utilized in the solution of the container feeder service network design problem have not previously been employed to solve a service network design problem and have in fact inspired later work on telecommunication network design (Rocha et al., 2009). Furthermore, the model resulting from the decomposition allows for a rich model of the service network design problem offering new levels of detail compared to previous approaches to this problem. Both problems, the container feeder service network design

and the network transition problem, are the result of collaboration with a major feeder operator and as such represent real problems that carriers face on a regular basis. The applicability of the proposed solution methods, however, extends beyond that of feeder networks and indeed beyond maritime applications.

Finally, the solution approach adopted for the network transition problem led to the development of a general parallel large neighborhood search framework capable of solving a multitude of different problems. The framework has currently been used to solve routing and scheduling problems as well as a variation of a facility location problem.

1.4 Thesis Overview

The following chapter will introduce the liner container shipping problem in further details emphasizing the connection with current planning processes executed by most liner carriers and with specific focus on processes related to the design of the service network. This chapter will also provide a discussion of current and future trends in the liner shipping industry with the purpose of determining the combined set of requirements imposed on the service network design.

Next, chapter 3 will contain a selected review of the literature relevant to the liner service network design providing insights into previous and current approaches to solving the liner service network design problem. The review will provide valuable input to chapter 4 which will discuss the modeling and solving of the service network design problem in the context of liner container shipping and also provide relevant parallels to aspects of more general network design problems. This chapter will also formalize the constraints imposed on the container feeder service network design problem.

The presentation in chapter 4 serves as an appropriate entry point to the two papers included in this thesis which treat the container feeder SND (chapter 5) and network transition (chapter 6) problems respectively. The papers are self-contained (including independent bibliographies) but reading chapter 4 prior to engaging in the papers will allow the reader to view the problems in a more broad perspective. Additional comments on the two papers are provided in chapter 7. Although chapters 2 and 4 provide an introduction to the SND and network transition problems, it is recommended to read the two papers before proceeding to read chapter 7. Finally, concluding remarks are provided in chapter 8.

The papers included in this thesis are:

Paper 1: *“Service Network Design in a Liner Container Feeder Application”*

The paper presents a service network design problem with cyclic scheduling and asset balancing requirements. The problem derives from a specific case in a liner container feeder application. By exploiting certain structural properties of the problem a solution approach based on a decomposition into two sub-problems dealing with route design and packing respectively is developed. The solution approach uses dual estimation in the calculation of new routes and utilizes an iteratively augmented central route pool to speed up convergence. Packing sub-problems determine feasible assignments of freight to individual vessels. These sub-problems have a knapsack structure and are relatively easily solved due to the reduction in size facilitated by the decomposition approach.

Computational experiments based on a series of cases adapted from real-world data show the feasibility of the algorithm with respect to solving realistic problem sizes. Furthermore, experiments suggest that the solution approach is capable of producing network designs with an overall high quality measured through e.g., capacity utilization.

Paper 2: *“Network Transition in a Liner Container Shipping Application”*

Closely related to the problem discussed in Paper 1 is the problem of transitioning between different service network designs. The paper proposes a parallel cooperative large neighborhood search heuristic to solve a network transition problem. The problem shares many characteristics with the pickup and delivery problem, but extends this with additional features specific to the network transition problem. This includes multiple time windows and delivery locations for commodities and routes that may start and end at different locations. Adopting a revenue-based perspective and allowing multiple delivery locations translates into the option of transshipping cargo.

The solution approach provides a relatively simple framework capable of solving a wide range of routing and scheduling problems. Specifically, the framework is well suited to solve tightly constrained problems where even determining a feasible solution can be difficult.

CHAPTER 2

Liner Container Shipping

The evolution of container shipping from its early days of industry resistance and slow adoption to its current role as a central component of modern globalized supply chains has been both dramatic and exciting.

When Malcom McLean launched the Ideal-X in April 1956, no one realized the impact this small vessel, carrying only 58 35-foot containers from Newark to Houston, would end up having on the global transport and manufacturing industry. It would take another decade of trial and error development of containers, vessels, and port equipment as well as strong union and port resistance before a standardized container saw international use. What followed was a dramatic growth in international trade fueled first of all by the decrease in the cost of transport realized through more efficient handling.

To fully understand the merits and drivers of containerized cargo transport, it is beneficial to first consider the situation as it was before the era of containerization. Figure 2.1 illustrates the typical flow of itemized cargo from the shipper to the consignee in a system with and without containerization respectively. Dark (red) shapes indicate the positions where manual handling of individual cargo items is required. From this figure it is clear that a significant amount of manual handling is involved in the transport of the cargo when containers are not used and that no significant consolidation occurs or is even feasible anywhere in the

network except during the main ocean haul. Transshipment of cargo is not economically feasible in such a system. Apart from the very high amount of time spent handling and stowing individual cargo items there are several other issues associated with this system. Manual handling means high labor requirements which adds significantly to the total cost of transport and also adds concerns regarding the security and integrity of the cargo items (Levinson, 2006, chap. 2). The cost of transportation of non-containerized cargo in 1960, could be as high as 25% of the total cost a product, with port related costs representing almost 50% of the total cost of transport (Levinson, 2006, chap. 1).

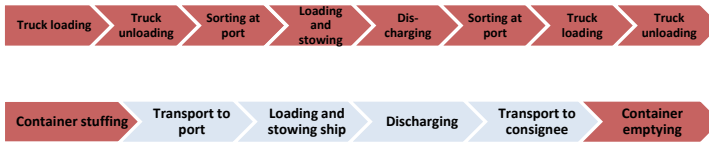


Figure 2.1: Manual handling (red shapes) of individual cargo items in a supply chain with and without containerization (adopted from Weldon (1958)).

Some early adopters of the new technology decided on hybrid solutions with break-bulk capacity below deck and space for containers above deck. This approach largely defeated the benefits of containerizing cargo, i.e., the quick turnaround of ships and reduction of cargo handling costs at ports. However, it did allow for some degree of more rapid cargo transshipment thus facilitating a move toward increasing consolidation and new route options.

From its early outset, *time and cost* have been two key driving factors in the development of container shipping. It was time that originally prompted Malcom McLean to introduce dedicated container ships in an attempt to reduce the duration of port stays required to load/unload a vessel carrying break-bulk, which in turn meant a dramatic decrease in turnaround time and thus more efficient utilization of the expensive assets that ships are. These improvements allowed for the transfer of derived benefits to the shippers in the form of reduced transport costs and reduced transit time. Taken in the perspective of modern day end-to-end supply chains, the derived benefits propagate throughout the entire chain ultimately leading to increased value and benefits for both end customers/consumers as well as manufacturers and suppliers. Early in the supply chain, suppliers can reduce their inventory and improve competitiveness through lean concepts and at the other end of the chain,

costumers/consumers get cheaper goods in a larger selection.

The container shipping and manufacturing industry has evolved significantly from the early days where the main focus was reducing transport costs and transit and turnaround times. As the level of integration in supply chains and intermodal transport systems increases, so does the level of complexity in the planning problems that liner carriers face. Especially in relation to the design and management of their service networks. To better understand the sources of the complexity, the remainder of this chapter will discuss some of the key features and challenges characterizing liner container service network design.

The following section will highlight some of the current and predicted future trends in liner shipping and comment on the potential impact for the design of liner service networks. Next, section 2.2 will discuss in greater detail the various planning processes related to the design and management of a liner service network. Emphasis will be on processes relating to the design of the network while processes dealing with more short-term maintenance and operation of the network (e.g. revenue management) will only be discussed briefly.

2.1 Current and Future Trends

Although cost and time remain the primary competitive parameters for liner carriers and determinants of success, customers are increasingly demanding more of the carriers than simple freight forwarding services.

In a point-of-view analysis conducted by IBM (Hingorani et al., 2005) for the container shipping industry several key trends were highlighted. Briefly summarized, these include:

- Strategic clarity
- Specialization (divergence of service)
- Integration of the supply chain (convergence of businesses)
 - Reliability
 - Visibility
 - Security
- End-to-end customer service

While the liner container shipping market has changed dramatically since the analysis was published as a result of the global economical crisis, many of the points remain valid and are expected to emerge as trends in the freight forwarding market, albeit delayed by the current economic recession.

In terms of strategic focus, three core target customer segments were highlighted in Hingorani et al. (2005) defined by their primary freight forwarding requirements; low cost, extended service, and value through integration. A carrier adopting the *low cost* strategy aims at providing the cheapest port to port service with high levels of standardization and service/schedule conformance. One of the problems with the adoption of the lowest-cost strategy is the threat of new players entering the market as customer loyalty in this segment is presumably quite low. At a cost of increasing requirements to maturity and business complexity, carriers may adopt an *extended service* strategy. Under this strategy, door to door service is the primary focus with success tied to the ability to offer reliable services, increased levels of information, and some level of customized products. Finally, a strategy of providing *value through integration* requires a strong focus on customer requirements, tight supply chain integration and a high level of availability of real-time information. Investments in customer relationships are required by rewarded by higher loyalty.

Central to all of the above three strategies is the ability to operate at the lowest possible cost within the chosen segment. Lack of strategic focus or inefficient operation of the service network will mean that stronger actors will take over the market. Similarly, to reduce the risk of performance instability due to highly manual planning processes and to maintain competitiveness, well-defined methods and systems to manage the service network must be implemented. Such systems must take into account the cross-network impact of business decisions, something which is only captured to a very limited extent by current practices.

In line with traditional thinking in supply chain management (see e.g., Stadtler and Kilger (2005)), the ultimate goal of carriers seeking to become an integrated supply chain logistic services provider is a deep integration with the businesses of the customers and tight coordination of freight, information, and financial flows. Traditionally, shippers have shopped around to secure the best value (perceived as service to price ratio) as well as spread out their freight across multiple carriers to reduce risk and dependency on a single carrier. Future relationships, however, are based on a consolidation of logistics service providers to enable closer integration of the supply chain. Such an integration not only requires a closer collaboration between customers and their portfolio of logistics service providers. In order to achieve the close integration required, investments are necessary both in terms of process alignment and relationship building. To justify such investments from both shippers and logistics service providers, more long-term relationships are required which explains the need for a concentration in the number of providers a single

customer will deal with (Levinson, 2006). The end result for the liner carriers is that fewer and fewer opportunities to secure business will arise as customers are increasingly engaged in long-term commitments.

Several capabilities precondition the move of a carrier toward a strategy of becoming an integrated supply chain service provider including reliability, visibility, and security of shipments. In particular, the demand for increased visibility has been driven by express package carriers whose core business is built on high quality shipment tracking information. Combined with modern communication and information technology (e.g., the Internet) customers have become accustomed to the availability of instant and accurate information and such expectations will be increasingly directed toward all logistics service providers, including liner carriers. Customers will evaluate their logistics service providers based on level of visibility and it becomes an important factor in building trust. Together, reliability, visibility, and security allow inventories to be reduced as less safety stock is required.

Overall, the future success of liner carriers rely on their ability to adopt a global view of their business which also means a departure from the traditional one-sided focus on asset utilization. Service reliability becomes central to securing business.

In the past few years, a new dimension in the form of environmental awareness has emerged further adding to the complexity and conflicting objectives of liner shipping planning problems. On one hand, environmental consciousness means investing in more efficient ships which will allow for a reduction in fuel consumption and result in cost savings. On the other hand, investing in new vessels is a costly and long term decision which means that during the fleet upgrade period, saving fuel may mean letting existing vessels sail at a speed lower than the nominal speed (slow steaming). This inevitably results in increased transit times and without fleet expansions also in reduced service frequencies. This obviously conflicts with the previously discussed goals of increasing customer service although it may increase schedule reliability through the occasional use of speed increase to reduce possible delays. This is also in sharp contrast to the otherwise pervasive focus on time and service highlighted above. However, as cost is the dominating competitive factor in liner container shipping, resulting savings may ultimately impact competitiveness in a positive way. Furthermore, speed of service remains a competitive parameter which is evaluated against the rest of the market and the negative impact on market share is reduced if it becomes the market norm to reduce speed. Finally, simply pursuing a “green” environmentally friendly image may just give the necessary competitive

advantage in a market with many similar services. An example of this strategy is the recent release of a CO₂-calculator by a major European feeder operator with the purpose of differentiating their freight forwarding services from e.g., truck or rail based transport modes by emphasizing the lower carbon-emission per tonne-mile, Hansen (2009). Ultimately, if the new requirements to environmental responsibility are met, the environmental focus can lead to a revitalization of the shipping industry through green competitiveness with other modes of freight transportation (Murray, 2007) and possibly release some of the current pressure on freight rates.

To set everything into perspective, many of the trends described above are currently not at the forefront of the focus for liner operators. The 2008 credit crisis and the following fall in consumer confidence and manufacturing has lead to dramatic decreases in container freight volume. Coupled with the large number of new vessels on order and delivered during the past year, the overall market is struggling to survive decreasing freight rates and a catastrophic over-capacity in the market. Efforts to counter the over-capacity include slow steaming with the fortunate side effect that fuel consumption decreases. Carriers are turning their focus inward with the objective of optimizing their organization. It could be argued that this is the right time for the deployment of advanced planning tools to help optimize service networks and this may also be the result in the coming years. Regardless, work on modeling and solving liner service network design problems should take the above aspects into consideration to be considered realistic when the market recovers. In this respect, research still faces several challenges as will become evident below and more concrete in chapter 4.

2.2 Planning Levels, Decisions and Objectives

Mapping the planning processes, decisions and objectives that are involved at the different levels of planning and execution of a liner service network provides a valuable input to the later modeling of the liner SND problem. To structure and break down the many processes associated with liner service network operation and to facilitate comparisons with related application areas the following sections will categorize each major planning process according to the traditional planning levels; strategic, tactical, and operational decisions. Associating a time horizon with each of these three levels of planning will aid the classification of processes.

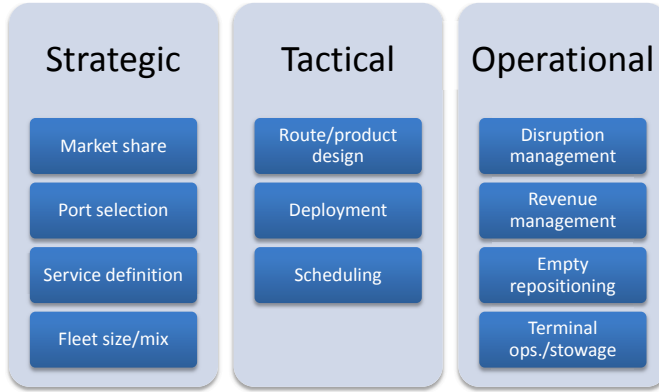


Figure 2.2: Overview of decisions and problems faced at each of the traditional planning levels; strategic, tactical, and operational. The listed tactical problems combine to form the service network design problem.

Figure 2.2 provides an selected overview of the major relevant planning processes at the three planning levels. Of particular interest are the tactical planning processes shown as these all combine to form what will be referred to as the service network design problem in the remainder of this work. It is relevant to note that each of these processes may in fact consist of several sub-processes and that the many additional processes and problems exist that are not represented in this figure. Additional problems can be found in Christiansen et al. (2007, sec. 2) together with references treating these problems in further detail.

Although the different planning processes will be presented as belonging to a certain planning level it is important to remember that many processes may span multiple levels and some processes are repeated at different levels with different degrees of detail. Planning is a continuous process in which decisions at one level will impact decisions at other levels. Higher level decisions set general policies and impose constraints on processes at lower levels and conversely these provide feedback and input to processes at higher levels in the form of financial and operational performance.

2.2.1 Strategic

Strategic planning refers to decisions made on a medium to long term horizon which in the case of deep sea liners corresponds to 1-5 years and for feeder liners corresponds to 1-3 years with some decisions (e.g., fleet

and terminal related) extending a further 5-10 years into the future. At this level of planning, knowledge about the future is relatively limited and typically associated with a high degree of uncertainty. Thus, decisions are made at higher levels based on aggregated information such as expectations to overall market development (e.g., refrigerated freight is going to increase, specifically out of a particular region), long term service goals/strategy, and expectations to market share. These decisions require deep business insight and are difficult to model reliably within the framework of operations research and thus go beyond the scope of this work. However, they serve as important input to the specific planning problems¹ addressed in further detail in this work by defining many of the parameters of these.

A fundamental exercise at the strategic planning level concerns the definition of both long and medium term *market strategy*. Identifying market potential, trends, and developments is the first step in this process but is complicated by the fact that the liner container shipping market is highly competitive and consists of a large number of actors. Together with an analysis of the competitive environment and a clear definition of the desired market position, necessary steps to achieve this position can be defined. Although partially outdated, Davies (1986) contains an interesting discussion and analysis of some of the mechanisms that govern competition and market conditions in liner shipping. Depending on the size of the carrier, differentiation and positioning in the market can be achieved through e.g., specialization or global coverage (with reference to section 2.1). Both of these strategies relate to service and while there are other dimensions in defining the market, service is probably one of the most important.

The overall *service strategy* will be defined by general expectations to the development of the market for containerized freight as well as the future competitive environment. The service strategy must at a high level define the major trade lanes that are offered as well as the more detailed aspects relating to frequency of service at major ports and transit time between regions. Again, multiple dimensions exist for several of these aspects. The competitive environment, for example, may be partially defined by liner conferences governed by rules under which the service network can be operated. Or the conferences may represent the external environment against which the service strategy will be measured. In such cases, service frequency rather than price may become the primary competitive parameter. A feedback also exists between the service strategy and the market expectations as it is generally accepted that demand

¹A planning problem may cover several planning processes in an integrated fashion.

for freight forwarding services will depend on the level of service that is offered. While this may not be prevailing at the overall level, it certainly becomes the case when more detailed plans are developed. E.g., it may be difficult to secure a satisfactory market with a bi-weekly service out of a certain port if a competitor is offering a weekly service.

Closely related to the service strategy is the *selection and location of ports* in the service network. Selection of ports will ultimately determine the markets that can be approached and the transport services that can be offered. The cost of freight handling at a given port will be an important determinant in the selection of ports. Aside from contracting with existing ports, some carriers may also be engaged in terminal activities. This means that at the very long term planning, analysis may have identified emerging markets in which a combined terminal development and liner presence is deemed promising. Such decisions will require the interaction of processes across several business units or companies to fully evaluate opportunities.

Finally, decisions regarding *fleet size and composition* are made at the strategic level as significant investments are involved in the purchase or chartering of vessels and building a new vessel can take several years to plan and execute. Fleet size and composition choices will be highly dependent on the overall service strategy as well as the forecast market share and planned network. On the other hand, network and service strategy will also depend on the technological possibilities of potential fleet expansions as well as costs. It will only make sense to expand the fleet if it supports increased revenues and/or improved market position. Thus, there is a complex interplay between defining the expansion and procurement strategy for the fleet and defining the service strategy. A situation where all the strategic planning problems come together is in merger and acquisition situations where the fleet, network, services, and market share may expand significantly over a short period of time. It is also in such situations that management faces very complex planning problems in trying to integrate the service networks of multiple entities.

Fleet size and composition will be addressed further in chapter 4 while the processes and challenges related to market identification, definition of service strategy and port selection will be assumed available as input for the remaining planning processes.

2.2.2 Tactical

Tactical planning focuses on events from two-three months into the future and up to one year and encompasses most of the planning processes that

will relate to what is referred to as a service network design in this work. For this reason, understanding the tactical planning processes is particularly useful to enable a rich treatment of the subject of modeling and solving liner service network design problems. Processes at the tactical level integrate all the strategic decisions to ultimately form a schedule that is the foundation of securing revenue. Most of the following applies to both short sea and deep sea shipping, but when significant differences exist, these will be noted.

The overall process of defining a schedule for a fleet of vessels to meet the requirements and restrictions to the service network established at the strategic planning level may be broken down into three sub-processes; 1. Route design (port sequencing), 2. Vessel deployment, and 3. Timing (scheduling).

Route design consists of the construction of a series of port sequences specifying how vessels are to move in the service network. The overall objective of the route design is to meet forecast demand using the available fleet at the lowest cost. The design of routes is a very complicated process as it must take into account requirements to service level including transit times and frequencies as well as evaluate alternative consolidation and routing options. The combined set of routes determines which final products that can be offered to customers. Although the port sequencing process may be executed independently of the scheduling process, there is always an underlying understanding of when certain ports should be visited governing the decisions. Fleet characteristics will impact route design with respect to topology choices. Section 2.3.3 will provide additional details on the impact of different route topologies.

Vessel deployment is concerned with the allocation of specific vessels to routes with the objective of maximizing the utilization of the fleet capacity. Deployment takes into account the expectations to freight volume on the individual routes as well as possible restrictions (physical or regulatory) of the types of vessels that can operate a certain route. Additionally, deployment determines the frequency of service offered on a particular route through the number of vessels assigned to that route. Deployment options are usually restricted by more long term decisions regarding fleet size and composition, but in certain situations, the fleet can be temporarily expanded through short term chartering. In these situations, deployment must take into account possible alternative charter options (see e.g. Rana and Vickson, 1988) in addition to the composition of the active fleet. Re-deployment of vessels on active schedules is not uncommon which means that the deployment process may be executed without changing the existing routes and schedule in any significant way.

Finally, *scheduling* is the concretization of the routes and fleet deployment into a combined schedule obtained by assigning specific timings to the port calls on individual routes. This process must respect any time related restrictions on ports such as closed or non-working periods and berthing time windows as well as take into account specific demand and availability patterns. Furthermore, scheduling must ensure the synchronization of routes that combine to provide products requiring freight transshipment. Since demand is generally dependent on the level of service offered (e.g. transit time) as well as the weekdays of calls, these are also aspects that must be handled in the process of creating the schedule. Although only a relatively limited degree of freedom is available in the scheduling assuming the routes are already determined, changing cruising speeds can extend or reduce the total duration of a route. Traditionally, this option has only been used to make up for lost time, but recently, whole route durations have been extended by slow steaming the assigned vessels to reduce fuel expenses. Notteboom and Vernimmen (2009) contains a discussion of some of the consequences of high fuel costs on the design of liner service networks and Bendall and Stent (1999) evaluates the economical impact of deploying fast container ships in long haul feeder services.

For deep sea liner operators, schedules are typically published six months in advance and usually see only small adjustments during the period of operation. For short-sea operators schedules are usually defined to enable efficient synchronization with the schedules of deep sea operators at the interfacing ports (generally hubs) but may change as frequently as every two to three months with deep sea scheduled calls at interface ports as pivot or fix-points.

Together, routing, scheduling and deployment represent the planning processes addressed in the service network design problem possibly integrated with fleet size and composition decisions in certain applications where this is feasible, e.g., with chartered fleets. Due to the high degree of interdependency between these different processes and the individual decisions within each process, the service network design problem is highly complex. Modification of a single route may have consequences for several other routes which is also why changes to service networks are mostly gradual and typically isolated to a few routes.

On the border between tactical and operational planning lies the network transition or network deployment problem which is treated in further detail in chapter 6. This problem provides the link between the network related planning conducted at the tactical level and the actual operational execution and realization of the decisions made.

2.2.3 Operational

Operational planning processes deal with decisions on the more immediate time horizon and up to a few months. Processes at this planning level are typically quite detailed dealing with individual assets such as a vessel, a booking, or a customer.

One of the vessel related processes is stowage planning. The objective of this problem is to position containers on a container ship such as to respect stability requirements as well as hazardous materials risks while taking into account the sequence in which containers are to be discharged to minimize the need for temporary repositioning. Efficient stowage of a vessel can significantly impact the cost of port visits as well as the time in port required to handle freight.

Although a large fraction of the total volume handled by a single carrier may be under longer term contracts, revenue management becomes increasingly important in a highly price competitive environment where further cost reductions are hard to realize. Revenue management processes deal with the question of accepting or rejecting cargo based on profitability relative to the available capacity, other offerings, and other cargo. On highly utilized routes, a revenue maximizing policy may involve rejecting a cargo to allow capacity for a later and more profitable cargo. Not much work has been published on revenue management in maritime applications but the success of employing revenue and pricing management in the airline industry may change this in the future. Lee et al. (2007) develops a heuristic for establishing the revenue maximizing cargo acceptance policy for a single leg in a container shipping application. This work takes into account the mixed nature of demand consisting of spot and contractual cargo.

Finally, a very significant and increasing problem associated with the operation of a liner service network is that of reverse logistics for empty containers, referred to as empty repositioning. The size of the problem is evidenced by recent trade statistics (Transmodal, 2009a,b) showing a 1:2.7 ratio between the freight from the Far East to Europe and Europe to the Far East and as much as 1:3 for the Far East to North America trade. This highly imbalanced trade means that a significant number of empty containers must be shipped to the Far East to satisfy the export requirements out of this region. As empty containers are typically not profit bearing², still consume capacity, and incur terminal handling costs, repositioning is a necessary evil for the continued operation of the service

²Carrying empty containers may be a profit bearing activity for some feeder operators.

network. For this reason, processes governing empty repositioning are very important to the success of a carrier as it not only ensures the timely availability of equipment (containers) but also provide feedback to processes dealing with product pricing.

A multitude of additional planning processes are carried out at the operational level including decisions regarding vessel bunkering (when and where to refuel), terminal operations (see Crainic and Kim, 2007), environmental routing aimed at determining the most economical routes based on weather and ocean currents, maintenance routing determining when and where to take a vessel out of the fleet for maintenance, and disruption management concerned with unforeseen changes in the operating environment causing changes to the schedule. In particular, this latter issue is very important as many uncontrolled factors impact the successful execution of a planned schedule which again is at the core of a reliable liner network supplying high service levels and supporting competitiveness. Section 2.5 will discuss some aspects of network disruptions, disruption management, and the consequences this may have for the SND.

Ideally, a model for the design of a liner service network should capture most of the planning problems discussed so far in an integrated fashion. To this end, some work (e.g., Gendreau et al., 2006) has been done integrating routing and stowage decisions although for a vehicle routing case where the route is not fixed. However, differences in the time horizon for the required execution and availability of information makes such an integration both impractical and problematic in a liner shipping context. In the above stowage example, schedules need to be published several months in advance of knowledge about attributes (weight, size, classification) of individual pieces of cargo required for stowage becoming available. Thus, it seems sensible and reasonable to structure attempts at modeling and solving the liner SND problem to capture planning processes and problems at the same overall level; in this case the tactical level with some processes stretching into the strategic planning. Realizing this approach in a set of concrete models is discussed further in chapter 4.

2.3 Service Network Components

While the processes described above serve to divide and detail the overall boundaries and goals of the processes that combine to form the service network design problem, they convey no detailed information about the

more specific properties that govern the service network.

The next sections will introduce in further detail the components of a service network: ports, vessels (container ships), rotations (routes operated by vessels), and demand. In the following, the term *service network* is restricted to denote only the sea-side of the full complex that is the end-to-end freight forwarding network. This means that hinterland logistics such as rail or trucking will only be briefly treated. However, parallels to related modes of transportation will be drawn when these are relevant.

2.3.1 Ports

Ports represent the nodes of the liner service network where all handling and transshipment of freight is carried out. Furthermore, ports provide the interface and gateway between land based logistics and the maritime network.

In general, ports will be classified as hubs and non-hubs (or outports) with hubs typically being major transit ports with a high volume of container throughput. Generally there is no specific pattern in the geographical distribution of ports, but several large ports may be located geographically close. It is relevant to note that close proximity from a geographical point does not necessarily imply that ports are close from a maritime perspective as land masses may separate these requiring significant sailing time to connect these.

Each port may have associated with it multiple terminals at which berthing can happen. Indeed, in many feeder networks, a single ship may need to call multiple terminals in the large hubs to load all outbound cargo. Certain ports will be closed for service during periods and many ports adopt differentiated pricing for container handling and berthing depending on the time during which service is performed. Thus, each port has multiple time windows associated each specifying an open/close status as well as handling and berthing costs. Surcharges may apply for service overtime which also means that time windows can sometimes be extended subject to payment of these surcharges. In general, a liner operator must negotiate fixed time windows for individual ports in advance during which arrival and berthing can occur. Depending on the port and specific terminal, missing such an agreed time window can result in a vessel being deferred until a new time slot becomes available. The time a vessel spends in port can vary depending on the number of cranes assigned to load/discharge containers and obviously the number of containers to be handled. High-efficiency terminals can handle approximately 40 containers per crane per hour.

Cargo handling beyond the shore side of a port is referred to as hinterland logistics and involve infrastructure such as rail and road networks. Hinterland logistics typically occur on a much less aggregated level in terms of volume and trucks and trains are responsible for the transport of cargo between the port and the origin/final destination. Although this work does not consider hinterland logistics, synchronization of schedules between ships and trains/trucks is important for the efficient operation of the end-to-end cargo transport. Depending on the nature of the collaboration between e.g., the liner and rail operator, synchronization requirements may be handled either through appropriate port time windows on the sea side or through the modification of train schedules after the liner schedules are defined.

2.3.2 Vessels

While vessels operating liner services exist in several variants such as RO-RO (roll-on/roll-off), reefer, and multi-purpose vessels capable of carrying both containerized freight and dry bulk, this work shall focus on fully cellular container ships. This type of ship will contain a compartmentalized hold (cells) to/from which containers can efficiently be loaded/discharged using specialized port-side equipment like gantry cranes. In this configuration, containers will be positioned on board ships in stacks up to around 15 containers high arranged in rows and tiers. Vessel volume capacity is measured in TEU (twenty-foot equivalent units) and range from 400 to more than 14,000 TEU with actual capacity dependent on container stowage and deadweight rating. Additional capacity restrictions apply in case of special requirements for the individual containers, such as refrigerated (reefer) containers requiring power plugs.

Generally, carriers will operate a very heterogeneous fleet consisting of vessels with different cost and physical (e.g., speed, capacity, and ice breaking capabilities) properties. For the largest vessels, the ports and canals that the vessel can navigate are restricted by physical constraints such as water depth and terminal equipment. This has so far lead to a classification of vessel size based on the largest possible size that can transit major canals e.g., panamax (Panama Canal), suezmax (Suez Canal), and malaccamax (Strait of Malacca). Depending on size, engine technology, and hull design, a vessel will be able to sail at speeds of approximately 20–25 knots with fuel consumption highly dependent on the speed. Vessel size will also determine the port and fairway fees that must be paid when calling a port.

Even the largest vessels only require a crew of between 10 and 20 persons to operate. In contrast to e.g., air crew and rail crew, the maritime sector is less regulated with respect to work schedules which also means that service network planning is generally not constrained by considerations to crew work scheduling.

A vessel will be either owned or chartered by the liner service operator. Owning a vessel means either purchasing an existing used vessel or ordering a new build. Prices for new builds depend on the size of the vessel but is in the order of \$100–120 million for a vessel with a capacity around 8,000 TEU meaning that investing in new vessels is very capital intensive. Charter rates have shown significant variations during the past few years with a mid 2009 price of around \$7,000/day for a 4,000 TEU vessel, down from more than \$30,000/day in the early/mid 2008. When chartered, contracts generally fall into four main types governing the division of responsibility for various aspects of operating a given vessel between the charterer and owner.

Time charter is the most common charter type in which the owner manages the vessel including the responsibility for crew, maintenance, and insurance. The charterer determines port call sequence and pays operational expenses such as fuel, canal fees, port charges, stevedoring, and other cargo related costs. A time charter may span anywhere from a few weeks to several years.

Slot charter contracts defines the division of capacity (slots) on a particular container vessel for a service or a set of voyages. This type of chartering is typically used in situations where multiple liner operators service common ports but do not have demand volumes large enough individually to fully utilize a vessel capacity while maintaining a high frequency of service. Through slot charters, operators can provide high frequency service on smaller trade lanes in a more efficient and economical way.

Bareboat charter and *demise* charter contracts transfer almost all responsibility of a vessel to the charterer. The charterer supplies crew and is responsible for maintaining the vessel. Charter periods typically fall in the order of years.

Voyage charter contracts govern the charter of a vessel to one or more specific voyages determined by a port pair. The vessel owner is responsible for all expenses related to operation including port charges and fuel. This type of charter contract is typically seen in tramp shipping.

In certain cases, a liner service operator will seek to reduce investments tied up in vessels by simply (time) chartering the entire fleet. This strat-

egy may be adopted by smaller operators that do not have the necessary funds or desire to purchase container vessels or larger operators that wish to maximize fleet flexibility and focus on the core business (freight forwarding). Under these conditions, the liner service operator will be referred to as a non-vessel operating common carrier (NVOCC) and function as a freight consolidator and forwarder. Adopting a NVOCC strategy introduces a new dimension into the medium- to long-term planning process which now also includes decisions about when to enter charter contracts and for how long. Depending on the contracts entered, this approach to fleet management can allow an operator to quickly adapt to changing conditions in the market. The impact of a NVOCC strategy on the service network design is discussed in Paper 1, chapter 5. Additionally, Rana and Vickson (1988) discusses the problem of chartering a single vessel and develops a complex model to aid deciding between charter alternatives.

2.3.3 Rotations and Products

Rotations comprise the backbone of the service network as they control the movement of vessels and thus the possible freight flows through the network. A *rotation* is a cyclic sequence of ports specifying the order in which ports are to be visited by vessels assigned to operate that particular rotation. In traditional service network terminology, rotations are also referred to as services with the term route used to denote the movement of a specific vessel. In the following, the term rotation is used to denote the generic specification of a visit sequence for a set of ports and to emphasize the requirement that vessel routes must start and end at the same port (form a closed cycle).

Topologies of individual rotations are diverse and depend on the nature of the demand and service level requirements for the ports covered by a concrete rotation. Thus, a service network may combine several different fundamental rotation topologies to offer different types of service. Out-and-back or pendulum rotations (Figure 2.3(c), bottom right), for example, require high freight volume to ensure satisfactory utilization and economy. This is also the primary reason that liner service networks are not operated as hub-spoke networks which is the predominant topology in e.g., airline networks. Regions with lower volumes of freight and minimum intra-region demand may be serviced by pure cycle rotations (Figure 2.3(c)) or variations such as butterfly rotations (Figure 2.3(b)) containing multiple calls to one or more ports possibly having higher volume. Finally, long haul services may be offered using what could be called conveyor belt rotations (Figure 2.3(a)) with major load/discharge

ports at the ends and visits to several large ports between these end points. Conveyor belt rotations typically require balanced demand in the main flow between the geographical end ports of the rotation and may not exist in the strict form illustrated. Variations and combinations of the different types of rotation topologies can be observed, but the topologies illustrated in Figure 2.3 represent the four major types found in liner service networks.

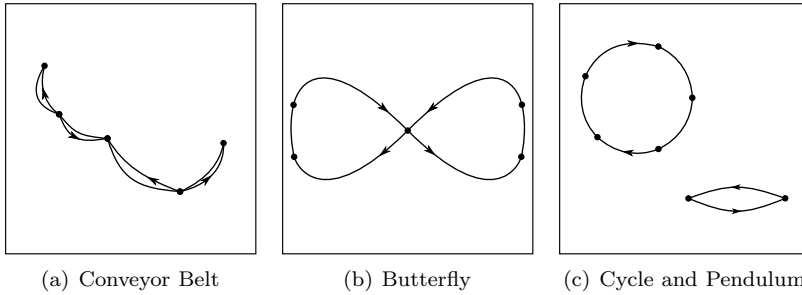


Figure 2.3: Examples of various liner network topologies. Arrows indicate travel direction and dots port visits. A complete liner network may consist of combinations of the above main types.

As noted above, close proximity of two ports in terms of direct distance does not imply short travel distance in a maritime setting as separating land masses may cause significant sailing time. Although not treated explicitly in this work, some carriers define rotations referred to as land bridges connecting ports using land based transport such as rail lines. These then function on terms similar to the ocean based rotations. This practice is particularly useful where reliable rail infrastructure is available and where the distance by sea is significantly longer than by land.

Where the rotations define the sequences in which vessels call ports (the flow of vessels) internally to a liner operator, *products* define the individual origin-destination transport services (freight flow) that are offered to customers (shippers). Products can be either direct or indirect services, the latter requiring the use of multiple rotations and thus transshipment of freight. Several products may exist between the same origin-destination pair, differentiated by transit time, cost, and schedule (departure date/time). Strategic planning usually defines a set of primary services that should be offered and the final service network design then reflects these requirements in concrete products and additionally offers derived products, typically requiring transshipments.

As previously mentioned, one of the primary benefits of containerizing

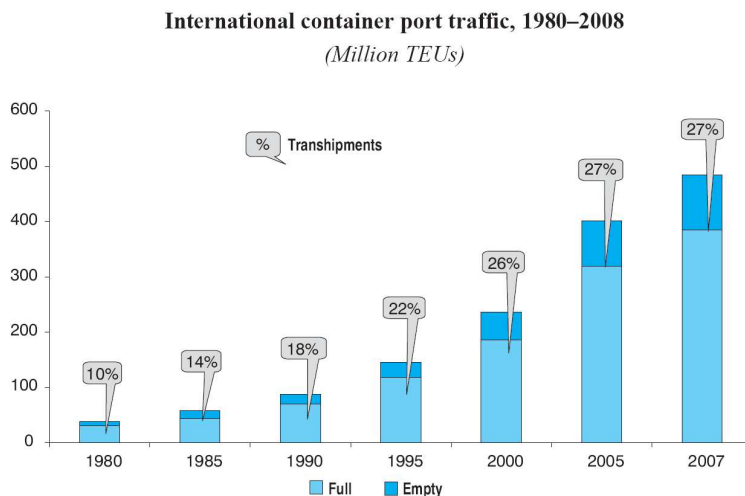


Figure 2.4: Development of full and empty container shares of the total port traffic as well as percentage of transshipments. Source: UNCTAD (2008).

cargo and operating a liner based network is the increased possibility of cargo consolidation. One result of extensive consolidation is the departure from offering direct service between all the ports of the service network in favor of a set of main services supplying direct service to only a small subset of the ports. However, with a known schedule and easy to handle cargo units, the option of cargo transshipment is introduced. Instead of having to offer direct service, a specific origin-destination port pair can now be serviced using a combination of multiple rotations through the transshipment of cargo at certain connection ports. With the correct structure of the service network, this leads to the offering of a new set of products which allows the liner operator to service a much larger set of origin-destination port pairs. Figure 2.4 illustrate this trend of increased use of transshipment based products.

2.3.4 Demand and Containers

Cargo is transported in containers which come in a large number of variants with the 20 (1 TEU) and 40 (1 FEU = 2 TEU) foot long variants probably being the most common. Generally, freight forwarding requests (demand) will be detailed as a set of origin-destination pairs with associated information about availability time and volume. Other attributes

such as weight or hazard classification may apply as well. Assuming mutually exclusive hinterlands, origin/destination can be interpreted as being specific sea ports.

In liner networks, a significant fraction of the total freight volume flowing through the system will be governed by longer contractual agreements. Such contracts represent a mutual obligation for the shipper to provide a minimum volume of freight and for the carrier to provide sufficient capacity for this volume. This, however, does not mean that freight volumes are always well known and static. Aside from variations in contractual freight, total freight volume is also determined by the availability of spot (ad hoc) freight. Furthermore, there are considerable seasonal fluctuations in demand e.g., with distinct peaks in the months before Christmas and significant reductions during Chinese New Year. Additional variation can be observed on weekly or monthly horizons necessitating the capture of demand dynamics in modeling approaches.

One of the major challenges when approaching liner SND is the availability of reliable demand forecasts. As most approaches to modeling the liner SND will likely use demand as the driving force in the design of the service network (see chapter 4), errors in forecasts can potentially lead to a service network unable to accommodate actual demand and secure revenue. Of particular concern is the use of forecasts based on historical demand patterns to re-design a service network as there will be a natural and strong correlation between the historical demand and the historical service network. This correlation will force the new service network in a direction that imitates the historical (existing) network in terms of both routing and scheduling. However, it is possible to mitigate the scheduling related issues by introducing larger time windows within which demand is available for pickup thus reducing the temporal correlation with the historical demand pattern.

Another issue related to the use of demand as the main driving parameter in the SND is the sensitivity of the network toward changes in demand patterns. Small scale experiments performed on simple compact formulations of the liner SND have shown that even small permutations of the overall demand pattern can lead to a significantly different service network design. Dealing with this issue is referred to as designing robust networks, the aspects of which will be discussed further in section 2.5.2. Ideally, approaches to modeling liner SND should be based on dynamic stochastic demand, rather than static deterministic demand. However, the resulting solution techniques become significantly more complicated as will the production of reliable forecasts as these must now include several additional dimensions that can be very difficult to estimate. Finally,

2.4 Differences in Short Sea and Deep Sea Operating Environments

even using dynamic stochastic demand does not capture the fact that demand depends on the level of service offered and accurately modeling this property again introduces significant complications into the models.

As mentioned earlier, reverse logistics related to empty containers is a significant challenge for container carriers. Although usually not modeled explicitly, empty container repositioning can, depending on the objective and structure of a certain model, be modeled as secondary demand and will thus compete with actual revenue contributing demand for space. One limitation of this approach, though, is that it does not accurately reflect the fact that empty container “demand” may be less time sensitive and can utilize secondary and slower products. Furthermore, empty demand is often defined by a fixed destination but may not necessarily have a specific origin thus complicating the modeling.

2.4 Differences in Short Sea and Deep Sea Operating Environments

The fundamental concepts, components, and planning challenges described above are relatively general and mostly apply to both deep sea carriers operating large service networks as well as to regional feeders with smaller fleets and service networks. Some differences between the operation of the two types of liner service networks have previously been highlighted, e.g., differences in the time horizons and frequency of execution of the various planning processes as well as characteristics of the fleet and the competitive environment. However, additional differences exist which impact the complexity of and approach to modeling and solving the liner SND for each of these two types of carriers. Table 2.1 provides a combined overview of the main differences between feeder and deep sea operating environments and service networks.

Deep sea carriers usually operate large networks to provide global service, secure high freight volumes, and achieve economy of scale and rely to a higher extent than short sea carriers on the transshipment of cargo. The presence of transshipment-based products is a complicating factor when modeling the liner SND problem. Simplifying assumptions ignoring these transshipment-based products are reasonable for some short sea carriers, thus greatly reducing the complexity of modeling and solving the feeder liner SND. Additional differences making the short sea liner SND more approachable typically include smaller vessel fleets and more flexibility in scheduling and fleet composition. In particular, the schedule flexibility means that short sea carriers may make frequent short term changes

	Short sea/Feeder	Deep sea
Geographical coverage	Limited, regional	Global, inter-regional
Volume	Small	Large to very large
Fleet size	Small to medium	Medium to large
Fleet ownership	Chartered or owned	Mainly owned
Vessel size	Small	Medium to very large
Capacity	Flexible	Fixed, short term
Schedule	Flexible	Fixed, six months
Consolidation strategy	Predominantly direct	Direct and trans. based

Table 2.1: Major differences between feeder and deep sea operating environments and service networks.

to their schedule and deployment in response to changes in operating conditions such as volume or the availability of spot freight.

Due to the global nature of deep sea carriers and the low cost per unit transported, there are no real viable alternatives to using liner container services when it comes to transporting break-bulk items. The competitive environment for short-sea carriers is very different, however, as short sea freight services may compete with rail or truck based transport modes for the main haul. Trucks are usually still required for the initial/final legs of the door-to-door journey, but substituting the main haul leg from truck to sea can help reduce road network congestion as well as emissions. An analysis of some of the challenges associated with such a substitution is provided by Paixão and Marlow (2002), proposing measures to promote and develop the use of short sea shipping including ensuring close integration between the different modes of transport.

2.5 Network Transition and Recovery

An underlying assumption for most of the discussions in this work is that model input is known and static, e.g., the forecast demand is constant and representative of the actually realized demand. In actual operations it is quite frequent that changes are made to the deployment and scheduling of the fleet as new information about demand patterns and disruptions becomes available. Due to the desire to operate the pre-published schedule, changes to the schedule are often sought isolated to a few vessels and introduced with the desire to maximize service while returning to the master schedule as quickly as possible. Emphasizing

the need for incorporating aspects of network recovery and robustness, a recent analysis (Vernimmen et al., 2007) highlights the impact of schedule deviations from the perspective of the shipper/consignee. In a world increasingly based on the principles of lean and just-in-time, unreliable schedules can have a great impact on the competitiveness of a carrier as well as its costumers.

The following sections will first address disruptions and possible actions to recovery from such disruptions and secondly outline possible approaches to mitigate the risks of suffering serious network disruptions. Disruptions are defined broadly to encompass any temporary event affecting the schedule in an immediate time frame. In the context of liner shipping, “immediate” is interpreted as being between 0–5 days from the event occurs.

2.5.1 Disruptions and Means of Recovery

Disruptions affecting the service network can usually be broadly classified as internal and external events. External events can be e.g., severe weather, strikes, congestion at ports, temporary port restrictions/closures, or volume changes. Internal events can be e.g., vessel break-downs, strikes, or deliberate decisions to change an upcoming port call or the destination of freight.

Disruptions, both internal and external, may lead to schedule delays, forced reroutings of both vessels and freight, and new, changed or dropped port calls. Delays may be compounded by the fact that the fleet operates 24 hours a day on tight schedules leaving little or no room for recovery. A missed time window can mean a delay until a new berthing slot can be found or a closed port which forces the vessel to either wait for the next available time window or bypass the port in question entirely. Furthermore, consolidation networks relying on transshipment of freight between vessels are particularly susceptible to disruption propagation. A delay in one vessel may cause delays to other vessels through cargo transshipments or require re-routing of the affected cargo.

Means to recover from a disruption include ignoring the event or making usually isolated changes to the schedule. Schedule changes can be broken down into a sequence of the following fundamental actions: move, omit, insert, and swap (basically two moves). A *move* involves moving a port call from one vessel to another, moving the port call in the visit sequence, or simply moving it in time. Alternatively, a port call can be entirely *omitted* and thus deleted from the schedule or if the disruption consists of a new freight opportunity, it may be desirable to *insert* an additional

port call. Finally, recovery from disruptions such as delays or changes in overall freight volume can be achieved by *swapping* two or more port calls between two vessels. This last action is most often used in short-sea shipping where vessels may have many schedule intersections.

2.5.2 Robust Service Networks

The concept of robustness relates to two main aspects of service networks; operation and design. Of the two types, robustness toward disruptions to the operation of the network has received the most attention in literature (see chapter 3). One reason for this might be that in the current planning practices of carriers (see section 2.2), the operational robustness is a very real problem that is dealt with on a regular basis whereas the predominantly manual network planning makes design robustness less evident. Essentially, *operational robustness* is some measure of the resilience that the network shows toward disruptions of some specified scale, i.e., disruptions that only affect a small part of the network. Building robust service networks means that tolerance toward possible disruptions must be introduced, e.g., in the form of added time slack in the schedule to make up lost time due to delays. In general, a service network can be considered robust if it provides sufficiently for the execution of one or more of the recovery actions outlined above, i.e., move, omit, insert, or swap port calls. An example of a network that is swap-robust is one in which schedules see frequent crossings, i.e., where two or more vessels call the same port within a short period of time (days) and cross again later in their schedule. If vessels have different capacities, such a situation will allow for a temporary increase of capacity on certain port calls while reducing it on others. While this scenario may seem artificial, it is actually quite common to perform such swaps in short-sea shipping in response to temporary changes in demand volume. Although some work has been done on designing robust service networks, approaches require either the solution of complex stochastic problems (see e.g. Birge and Louveaux, 1997) or incorporating specific properties of service network designs that are considered robust in the modeling. The latter approach currently seems the most viable and indeed some work based on this idea has been published for maritime scheduling (Christiansen and Fagerholt, 2002) and airline scheduling (Smith and Johnson, 2006). In the first work, robustness is attributed with avoiding risky arrivals which are arrivals close to the end of time windows and in the latter, a robustness attribute of an airline fleet assignment (station purity) is incorporated into the model and solution method. Additionally, Ageeva (2000) discuss various properties that define robust airline schedules, including the above proposed

swap option.

Design robustness has previously been briefly discussed in connection with demand forecast in section 2.3.4 but extends beyond demand to include essentially all input to the SND. The issue that arises in many approaches to modeling and solving SND problems is that the final designs are highly sensitive to even small changes in the input data. Changing the capacity or cruising speed of a single vessel or adding one additional commodity can result in an entirely different service network. Typically, changes are not isolated to a single vessel but tend to affect the whole network due to the many interdependencies that exist between the different components. Building models and solution methods to reduce the sensitivity toward changes in input is not an entirely simple task. It would require the introduction of some measure of deviation from an accepted baseline to evaluate whether or not a proposed solution to an SND problem has changed too much. Unfortunately, it is not obvious how such a baseline is to be established, what should constitute a deviation, and how deviations should be bounded. It may be argued that simply using the current service network as a baseline is a straight forward approach, but this may defeat the whole purpose of modeling and solving the SND problem to begin with, i.e., to optimize the service network relative to new conditions. Some work has been done in an airline context (see e.g. Klabjan et al., 2001) although the underlying problem is somewhat different from the liner SND. In a more general approach, Brown et al. (1997a,b) use the term persistence to denote design robustness and present a series of case studies from various industries to support the need for incorporating persistence into models to increase managerial acceptance of optimization based systems. This work will not address the issue of design robustness in further detail, but it is clearly an issue that should be addressed in future research. Until a satisfactory solution is found, it is expected that any system developed to design service networks will need to rely on skilled human interaction to evaluate the fitness of suggested network designs in relation to the business.

2.5.3 Network Transition

The previous sections briefly dealt with unplanned disruptions with respect to causes, schedule impact, and mitigating or recovery actions. However, planned changes to the active schedule occur several times a year as a result of the transition from one SND to another, possibly caused by a change in schedule, demand patterns or fleet composition. The reason for treating network transition together with the subject of disruption management and recovery is that all these processes share a

common fundamental structure. In fact, the concept of network transition, i.e., the problem of migrating part of or the entire fleet from one service network design (and associated schedule) to another design (and schedule), encompasses the challenges also faced in network recovery processes. It is simply a matter of defining the appropriate time horizon and involved set of vessels. Where a planned major network transition might take place over a period of weeks, a recovery might only take place over days or hours and only deal with a few vessels.

The network transition problem is discussed in further detail in Paper 2, chapter 6 where an algorithmic framework to solve this problem is also developed. Extensions of the network transition model and concepts to allow a carrier to move toward continuous dynamic network planning are discussed in section 7.2.2. Additionally, section 7.2.3 will interpret network recovery operations in the context of the network transition problem and highlight required extensions and limitations.

CHAPTER 3

Literature Review

Transportation planning and optimization is one of the classical subjects within operations research which has received considerable attention from both researchers and practitioners throughout the history of this academic discipline. And for good reasons. Transportation problems are diverse in nature and offer many interesting challenges both from a research and from a practical perspective. Additionally, transportation systems represent a significant part of most national and international infrastructures as well as playing an important role in the continued growth of the world economy.

The purpose of this section is to provide a selected review of published work on maritime liner network planning with a specific focus on works addressing service network design using mathematical programming and operations research methods. The presentation groups references based on the primary solution approach employed rather than the specific problem class. Historical developments, if any, are highlighted and compared to trends in related problem areas. Due to the relatively few works dealing with liner (container) based service network design, references to related application areas within transportation will also be provided when relevant. It is important to note that although the problems addressed in this thesis deal exclusively with liner container shipping, other forms of liner shipping do exist (e.g., passenger cruise Hersh and Ladany (1989)). However, container based liner service is the most common type of liner

based shipping owing to its flexible cargo handling and does present a series of challenges not often present in other forms of liner shipping.

It is not the goal of this section to provide an exhaustive review of literature on operations research applied to maritime routing and scheduling problems. Reviews have been conducted by Ronen (1983, 1993) and Christiansen et al. (2004, 2007) with the latest containing a comprehensive survey and description of a wide range of maritime planning problems including those related to the design of the service network. Nor is it the goal to develop a new taxonomy or classification scheme for maritime routing and scheduling. Ronen (1983) provides an early classification scheme that covers maritime freight transportation in general and Kjeldsen (2008) presents the first steps toward developing a scheme specifically for liner shipping problems.

Table 3.1 lists a number of publications dealing with maritime routing and scheduling based on a recent surveys by Ronen (1983, 1993), Christiansen et al. (2004, 2007) and supplemented with a few reference not present in these surveys. It is interesting to note that only 18 of the 136 published works from the three earliest surveys deal with liner based shipping in one form or another. It is important to note that although routing and scheduling in liner networks using operations research techniques has received relatively little attention, publications about policy, regulation, and economics of liner shipping are more numerous. These are however not included in this work and the reader is referred to Broeze (2002) for an entry point this area. Furthermore, literature on routing and scheduling in non-liner shipping applications is also more rich. However, as it is evident from the discussions in chapter 2, it is relevant to distinguish strictly between liner and non-liner related works as some of the assumptions in other problem areas in the maritime literature, e.g. tramp shipping, do generally not hold for liner based problems.

Despite showing an overall increase in attention, there is no clear connection between the subjects of the published works and historical events such as the rate crises in the mid 70's and mid 80's or emerging trends such as those described in section 2.1 (e.g. changing service requirements and integration of supply chains). Equally interesting is the observation that although there is a slight increase in the number of publications during the past decade, it does not match the explosive growth in the capacity of the global container vessel fleet. This is particularly surprising given the planning challenges that liner service operators face following recent years of expansion and consolidation (see section 2.1 for a discussion)

Reference	Method	Features/Type
Agarwal and Özlem (2008)	Col.gen., Benders, heur.	weekly schedule
Almogly and Levin (1970)	LP	Stochastic
Boffey et al. (1979)	LP, Heuristic	Decision support
Cho and Perakis (1996)	ILP	Fleet size/mix, deployment, routing
Claessens (1987)	LP, Heuristic	Route selec. and deployment
Datz (1968)	Analysis	Expr. for schedule eval.
Fagerholt (1999)	Col.gen.	Set partitioning
Fagerholt and Lindstad (2000)	Col.gen.	Gener. set covering
Fagerholt (2001)	Col.gen.	Set partitioning, soft TWs
Fagerholt (2004)	Col.gen.	Set covering
Hersh and Ladany (1989)	Dyn.prog.	Passenger cruise
Kydland (1969)	Simulation	Routing, scheduling, fleet size/mix
Lane et al. (1987)	IP, Route enum	Set partitioning
Olson et al. (1969)	Simulation	Medium-term pln., ad-hoc analysis
Perakis and Jaramillo (1991), Jaramillo and Perakis (1991)	LP	Deployment
Powell and Perakis (1997)	IP, Route enum	Service freq. driven
Rana and Vickson (1988)	Lagrangian relax., Benders	Non-linear, single-vessel charter
Rana and Vickson (1991)	Lagrangian relax., Benders	Non-linear, routing and scheduling
Shintani et al. (2007)	Genetic alg.	Routing, scheduling, integr. empty repos.
Sigurd et al. (2005)	Heur. col.gen.	Visit separation, max lead time
Weldon (1958)	Analysis	System design and analysis
Weldon (1959)	Simulation	Stochastic demand

Table 3.1: Overview of literature on routing and scheduling in liner container shipping.

3.1 Simulation Based Approaches

Some of the first published work on liner container shipping is Weldon (1958) and Weldon (1959). In the first paper Weldon sets up a system study to evaluate the profitability and feasibility of introducing some level of containerization of the general cargo on a specific trade lane at Matson Navigation Company. This study was carried out at a time where the concept of containerized cargo was still very new and a large part of the study is dedicated to determining properties of the relevant assets. This includes fleet size and composition and equally important and correlated, determination of container dimensions. The paper provides insights into the early decisions that liner service providers faced at the brink of paradigm shift as well as an analysis of the cost involved in the necessary technology change.

In the subsequent paper from 1959, Weldon builds on the findings of the first and develops a simulation model which provides planners with a tool to rapidly evaluate schedule performance under varying demand conditions. The core of the developed model is the demand generator which allows a planner to specify a set of scenario parameters (referred to as boundary conditions) and obtain a demand scenario under which a particular schedule is evaluated. Common for both papers Weldon (1958, 1959) is that they pioneered the use of systematic operations research techniques for a liner container shipping problem. However, none of the papers deal with any form of automated schedule generation, although this is mentioned as a possible extension in Weldon (1959).

Olson et al. (1969) to some extent continued the work started by Weldon at Matson Navigation Company. At this point, the company was already partially engaged in liner container shipping and Olson et al. developed a simulation model with the purpose of determining maximum profit schedules. The model takes into account initial vessel positions and loads as well as operational constraints dictated by the management. Using the model the authors simulated several 90-day scenarios for a limited set of tradelanes and found profit increases of up to two percent compared to the current plan.

Following the same general principles as Olson et al., Boffey et al. (1979) develops a decision support system consisting of two parts. The first part is a simple interactive program that performs consequence calculations based on routes input by a planner but does not contain any algorithms to suggest such routes. For the second part, the authors develop a simple greedy heuristic to iteratively select and sequence a set of ports into a route and employ an LP model to evaluate revenue of a given port

sequence (route) by determining optimal scheduling and cargo routing information for this sequence. The LP model incorporates transit time dependent demand. Based on evaluation of the interactive program at a liner operator and presentation of the heuristic the authors conclude that management “liked and understood” the interactive program, but showed less interest in the heuristic route generator. This supports the notion that sometimes the first step is to collect and present all relevant information in a structured manner. Although the approach adopted by Boffey et al. is simple from an OR perspective, the paper offers some interesting comments and observations about the liner container business many of which remain true today. Furthermore, it is possibly the first paper to model service dependent demand (later, Bendall and Stent (2001) develop a model with service dependent demand for a hub-spoke feeder service and propose a two stage approach to solving this model similar to a route-first schedule-second principle).

3.2 Decomposition/Relaxation Based Approaches

Rana and Vickson (1988) develops a non-linear mixed integer model for the problem of determining the profitability of chartering an additional vessel into a given network of ports with a known excess demand. The model determines the best sequence of port calls under the assumption that the start and end ports are given and that the vessel performs an outbound and inbound voyage (trip) visiting zero or more ports between these end points. Any port can be visited at most twice, once on the outbound and once on the inbound voyage, and an ordering of the ports is assumed restricting route topology. The model is linearized by solving multiple times with the complicating variable fixed at a set of discrete values. The model is partitioned into a cargo allocation sub-problem and a route design problem using Benders’ decomposition and Lagrangean relaxation. Computational experiments are performed on networks with between 10 and 20 ports and the Lagrangean based and a specially designed algorithm exploiting specific model properties are compared. The largest problems are reported to be solved in minutes. Later, Rana and Vickson (1991) extend the model of Rana and Vickson (1988) to capture multiple vessels and relax the assumption on fixed route end points by introducing additional variables. This adds additional non-linear constraints that are handled through variable fixing and decomposition. The route topology restrictions of the previous model remains. The model is solved using decomposition and Lagrangean relaxation as previously. The authors report solving problems with three ships and up to 20 ports

in approximately 3000 seconds for the largest problem.

Using a more traditional modeling approach, Agarwal and Özlem (2008) solve a ship scheduling and cargo routing problem using a variation of the classical capacitated multicommodity network design problem. The objective is profit maximization and the demand is assumed to have a weekly repeating pattern but all demand need not be carried. While the modeling approach allows for the transshipment of cargo between different routes, the cost of such transshipments is not captured. As the cost of transshipping containers is large compared to the total cost of transport (see section 2.3 for additional details), this exclusion can lead to solutions inappropriate for real-world implementation. To evaluate the effect of ignoring transshipment costs, the authors perform post-analysis on the solutions by re-flowing cargo with varying levels of transshipment costs and conclude that the volume of demand that can be profitably carried in the service network proposed by the model decreases by around 36% as the transshipment costs increase. The model is solved using an innovative combination of a two-phased Benders' decomposition combined with heuristic column generation using an iterative search to identify negative reduced cost cycles (routes). This method is compared with a simple greedy heuristic and an approach based on heuristic column generation for the non-decomposed model. Computational experiments are performed on synthetic test instances with up to 20 ports, 100 vessels, and 114 demands and results show that the two-phased Benders' decomposition approach outperforms the other two approaches both in terms of solution quality and execution time. The test instances of Agarwal and Özlem are quite different from those used in chapters 5 and 6 in terms of the balance between vessels and demands as well as demand patterns. Appendix B contains a more detailed discussion about issues related to data generation.

3.3 Column Generation Approaches

Fagerholt and Lindstad (2000) treats a problem where a set of oil exploration installations must be serviced from a central depot and the best fleet size and composition for providing this service must be determined. A weekly schedule is desired and each vessel can operate multiple shorter routes as long as the total duration of these does not exceed one week. Ignoring compatibility between the shorter routes served by individual vessels is possible since time windows at both depot and oil exploration installations are independent of weekday. However, since no correlation between routes is captured, the approach may potentially lead to solu-

tions where certain installations have all their visits grouped together in a short period of time. A two-phased approach is adopted in which candidate routes are generated in the first phase to serve as input for the second phase which solves a variation of a set covering problem. Fagerholt and Lindstad analyze a series of scenarios and conclude potential cost savings are in the order of 43%. Using a very similar approach, although without the fleet size/mix problem, Fagerholt (2004) addresses a feedering problem for the transport of freight from a set of production units to a central hub. Weekly schedules are determined for problems with up to 40 ports.

In a more rich context, Sigurd et al. (2005) treats a problem where a transport tender specifying service requirements for the transport of containers between a defined set of locations is given and the objective is to determine the least cost fleet size, composition and schedule that satisfies this tender. The problem introduces requirements to recurrence and separation between visit times to ports as well as allowed lead time from pickup to delivery of individual cargoes. Capturing these cross-route constraints is achieved by augmenting a set partitioning formulation with pre-defined legal visit patterns. The authors adopt a branch-and-price approach solving the route generating sub-problem using a tailored algorithm exploiting the special structure in the network; a remotely located hub serves as a vessel depot from which all vessels depart and return to before then end of the planning horizon. Assuming no cargo crosses the hub, Sigurd et al. use ideas from the vehicle routing literature and formulate the first phase of the two phased pricing sub-problem as a resource constrained shortest path problem. In addition to satisfying operational constraints, the authors impose a set of heuristic constraints to reduce the complexity of the sub-problem. Routes generated in phase I of the sub-problem are stored and combined into full-length routes covering the whole planning horizon in phase II. Ryan-Foster branching on consecutive pairs of visits augmented with a scoring system based on the relation between such pairs is used as branching strategy. A real-world case problem with 68 weekly cargoes in a network of 21 ports and ten possible ship types is solved and savings of 14.8% are reported when extending the planning horizon from one week to two weeks due to the increased flexibility in route compositions.

3.4 Related Literature

Liner SND problems share many structural properties with vehicle routing and scheduling problems. In particular, under appropriate assump-

tions, a simplified version of the liner SND can be modeled as a pickup and delivery problem (PDP) (see e.g. Toth and Vigo (2002), Berbeglia et al. (2007), Parragh et al. (2008a,b) for an introduction and recent surveys). Although these assumptions generally mean that a more strict structure is imposed on positioning of the fleet, the scheduling of individual routes, as well as the flow of freight, many of the techniques that have proven successful for the pickup and delivery problem can be adapted to liner SND problems. This is in particular the case for the network transition problem treated in chapter 6.

Two papers, Cortés et al. (2010) and Mitrović-Minić and Laporte (2006) consider a special class of pickup and delivery problems of particular interest referred to as the pickup and delivery problem with transshipments. In this problem, the PDP is extended with a set of fixed transshipment points at which requests can be transferred between vehicles. However, the presence of transshipment points further complicates the solution of the PDP as it introduces additional synchronization constraints. The largest problem considered by Cortés et al. contains six requests, two vehicles and one transition point and takes around two minutes to solve (although solution time increases with a factor of almost 250 times compared to a problem with half as many requests). Real-world instances of liner SND problems can contain hundreds of requests and multiple transshipment points suggesting that more work is needed before the proposed approach is applicable in real-world planning scenarios. Mitrović-Minić and Laporte report solving instances with up to 100 requests and four transshipment points using two-phased heuristic noting that computational time was small although without reporting any specific numbers.

Additional related problem classes include airline scheduling and fleet assignment, express package service network design and to some extent also certain aspects of train timetabling. Barnhart et al. (2003) provides an introduction and overview of several planning processes in the air transport industry including the above mentioned scheduling and fleet assignment problems. Interestingly, only a relatively limited amount of work is published on the subject of airline schedule design and the authors attribute this to the limitations in the currently available methods to capture the full scope of the schedule design problem. Noting that the current practise is to construct flight schedules manually, the authors also suggest that there is a growing trend towards adopting optimization based methods in this part of the planning problem as the methods improve.

Finally, though not strictly dealing with aspects of designing and main-

taining a service network from the perspective of operations research, Lawrence (1972) provides a more broad discussion of maritime transport with primary emphasis on freight transportation both from a technical and economical perspective. In a more recent work, Broeze (2002) focuses specifically on containerization and the impact it has had on the global economy and trade. A similar focus on economy is found in Levinson (2006) which also contains a colorful presentation of the history of containerization, from initial small experiments with container sizes and handling equipment including challenges relating to infrastructure and regulations (e.g., union resistance) to the current day global dependency on containerized freight transportation.

3.5 Summary

With an apparently ever increasing size of service networks as well as the complexity of the requirements for its design and operation, it is important to remember that what may be considered a realistic problem size today, may be considered a small problem when research is implemented and operationalized a few years from now. Currently, the growth rate of the size of real-world problems exceeds the progress in standard linear and integer programming solvers. Even employing advanced decomposition and solution techniques combined with utilization of problem specific properties, many of the above reviewed works suggest that solving larger real-world problems is associated with significant difficulties. Possibly even more limiting from a business perspective is the fact that only a few capture more than the basic structural properties of what is in reality much more rich liner service network design problems.

Perhaps, going forward, solution and modeling approaches should focus on understanding the current planning processes of liner carriers and seek new ways of thinking about problem decomposition, possibly inspired by these processes. Equally important for business adoption is the development of flexible and robust solution methods that can be adapted to new operating conditions while maintaining computational tractability. The next chapter seeks to provide some initial input to such developments through the analysis and comparisons of the merits of various modeling approaches for the SND problem.

CHAPTER 4

Modeling and Solving the Liner Container Shipping Service Network Design Problem

So far, the liner service network design problem (SND) has been addressed primarily from the perspective of the business processes that define and support it. The previous three chapters served to provide an introduction to and broad background on liner container shipping with specific focus on elements related to the design and maintenance of the service network. The purpose of this chapter is to extract and formalize the main properties and requirements imposed on the SND with the purpose of supporting an analysis and development of models and solution methods to support the SND related planning processes.

Discussions in this chapter will focus on the tactical planning problems according to the definition of the service network design problem in chapter 2. To summarize, the planning processes that must be captured by a liner SND model are the following (from section 2.2):

1. Fleet size and mix
2. Routing/product design

3. Scheduling
4. Fleet deployment

With the exception of item 1, all of the above processes are executed at the tactical planning level. Although typically carried out as independent processes, the quality of the final service network will benefit from an integrated approach capable of capturing the interaction between decisions made in the various processes. It is, however, crucial to the success of an integrated modeling approach to ensure consistency between input data, decision abstractions (variables), and the desired output. E.g., modeling revenue in a tactical SND model makes little sense since such information is probably not accurate or may not even be available at that level of planning. Thus, all elements of a model integrating multiple planning problems should support a common overall design goal and be consistent in the time horizon on which decisions occur and input becomes available. As it will become clear later in this chapter, fully integrated models are not without complications and problem decomposition may be necessary to make models computational tractable. Thus, the appropriate level of process integration in the modeling approach is determined by a compromise between computational tractability and the required quality of the service networks resulting from solving such a model.

While the processes may provide inspiration to possible modeling and problem decomposition approaches they convey no specific information about the more detailed requirements that a service network must satisfy. To aid the development and evaluation of model alternatives, a summary of the primary requirements and properties described in section 2.3 is provided below. The satisfaction of these will be discussed along with each of the models presented later in this chapter.

1. Multiple *intersecting* routes (rotations)
2. Routes are *cyclic* (i.e., design must balance)
3. Multiple *capacity* dimensions on vessels
4. Actual capacity may depend on route
5. Costs depend on vessel cruising speed
6. Vessel-port *compatibility* limits feasible routes
7. *Time windows* limit port visits
8. *Service level* requirements are imposed on ports
9. Freight can be *transshipped* between routes
10. Demand varies over the planning horizon
11. Demand may depend on service level
12. Cost components include: port fees, fairway dues, terminal handling, vessel charter, bunker.

One of the main features that differentiates the liner service network de-

sign problem from traditional vehicle routing and scheduling problems is the cyclic nature of the routes. This means, that a general model must be able to capture the consequences of demand flowing past the end of the planning horizon and into the beginning consuming capacity throughout. This apparently simple requirement complicates or voids the use of algorithms traditionally employed in vehicle routing and scheduling problems. Some works on liner SND, e.g. Sigurd et al. (2005), avoid the complexity of cyclic routes by making assumptions about how demand flows. In general, however, it is necessary to deal with the cyclic requirement when modeling liner SND problems. Another feature of the liner SND is the presence of the option of transshipping freight between routes. This adds significant complexity in the modeling of freight flows as well as requirements to synchronization between routes. Finally, although not specific for liner service networks, the fact that demand generally depends on the level of service offered represents a significant challenges in terms of modeling the SND problem as most approaches are based on demand driving the SND and do not include explicit feedback between service level and demand.

Given the many levels of decisions and multi-dimensional correlated requirements that a liner service provider faces in the design of their service network, the following sections address some of these decision problems through the development of a series of models. The models evolve from simple problem abstractions that cover only a very limited set of the constraints and requirements imposed on liner service networks to gradually capturing more complex representations of the problem. Where the first models capture only basic routing decisions and equipment balancing requirements, later models allow the implementation of very complex requirements to e.g. route topologies, load constraints, and service level requirements. Ultimately, any proposed model represents a compromise between computational complexity and meeting business requirements. Common for all the models is that it is a demand forecast that drives the design of the service network and that this demand is assumed known and static throughout the modeled planning horizon.

As the model complexity increases, there is a natural progression from simple dis-aggregate decisions toward increasing levels of decision integration. The main point of this chapter is that for some classes of service network design models, the more information that is contained in a single decision, represented formally by a decision variable, the better properties the problem gets from the point of view of solving this using e.g., standard mixed integer programming software. Although rarely stated explicitly, this fact is not new knowledge and has been implicitly suggested and exploited by the increasing number of works using solution

approaches based on the principles of complexity delegation through decomposition and variable aggregation (Armacost (2000), Armacost et al. (2002) is an example of the latter). In these approaches, it is the strong integer properties of the models that improve tractability when solved using standard polyhedral methods. As it turns out however, aggregating decision information leads to a series of other computational challenges that sometimes outweigh or entirely void the benefits gained by having better integer properties. Heuristics, on the other hand, are not as sensitive to these properties and currently represent the most viable way to addressing realistically sized network design problems. Regardless of the approach, the end conclusion remains the well known: network design problems are indeed very hard to solve (to optimality).

4.1 Multicommodity Capacitated Network Design

Network design has been the subject of intense research for more than five decades and is indeed one of the truly classical problems within operations research. And for good reason that is. Classical network design problems have many interesting applications in real-world problems. Furthermore, they are deceptively easy to formulate but very difficult to solve even in the most simple cases. Problems generally exhibit highly structured combinatorial properties that make them intriguing from a theoretical perspective.

Basically, two fundamental decisions are modeled in the multicommodity capacitated network design (MCND) problem; the flow of freight and the operation of assets. In the MCND, operation of assets can be interpreted as opening a service between two nodes of the network. Depending on the network representation, a node may simply correspond to a terminal or it may be part of a time-space network¹ and thus represent a terminal at a discrete point in time. Only in the time-space network representation does MCND allow for the modeling of scheduled services. Given the two decision components, the MCND problem then becomes that of determining the best deployment of a set of assets to satisfy demand while respecting capacity constraints. In this aspect, the MCND extends the classical multicommodity network flow problem (see e.g. Ahuja et al. (1993)) with the challenge of determining on which arcs of the network to provide service

¹Refer to page 79 for a brief description

In terms of modeling the liner SND, the multicommodity capacitated network design captures the following aspects:

- All demand must be met
- Demand varies over the planning horizon
- Servicing demand must respect asset capacities

It is clear that the MCND problem represents a very limited model of the liner SND specifically by the omission of asset balancing constraints and fleet size constraints. Additionally, although the MCND can model different capacities on different links of the network, the absence of a notion of individual vessels makes this less useful. However, despite its limitations, the model is a very useful tool in understanding some of the difficulties associated with modeling and solving network design problems. Furthermore, due to the fundamental concepts that are introduced by the MCND model, it serves as an invaluable introduction to the more rich models that follow later in this chapter.

4.1.1 Mathematical Model

Let $G = (\mathcal{N}, \mathcal{A})$ denote a graph with nodes \mathcal{N} and directed edges (arcs) \mathcal{A} . The graph can be a time-space representation of a service network, but the model does not make any such assumptions about the graph. The set of commodities flowing through the network is denoted \mathcal{K} . Each commodity $k \in \mathcal{K}$ is defined by an origin $O(k)$, a destination $D(k)$, and a volume (amount) d^k . Using standard notation, let $y_{ij} = 1$ if arc $(i, j) \in \mathcal{A}$ is open for service supplying a capacity of u_{ij} units and let $y_{ij} = 0$ otherwise. Furthermore, let x_{ij}^k denote the volume of commodity $k \in \mathcal{K}$ flowing (transported) on arc $(i, j) \in \mathcal{A}$. A fixed cost f_{ij} is associated with the opening of service on an arc and an additional cost of c_{ij}^k is incurred by the transport of one unit of commodity $k \in \mathcal{K}$ on this arc. Finally, let $b_{ij}^k = \min(d^k, u_{ij})$ and $d_i^k = d^k$ if $i = O(k)$, $d_i^k = -d^k$ if $i = D(k)$ and $d_i^k = 0$ otherwise. With these definitions, the basic capacitated multicommodity network design problem can be expressed as follows:

$$(\text{MCND}) \min \sum_{(i,j) \in \mathcal{A}} f_{ij} y_{ij} + \sum_{k \in \mathcal{K}} \sum_{(i,j) \in \mathcal{A}} c_{ij}^k x_{ij}^k \quad (4.1)$$

s. t.

$$\sum_{j: (i,j) \in \mathcal{A}} x_{ij}^k - \sum_{j: (j,i) \in \mathcal{A}} x_{ji}^k = d_i^k \quad \forall i \in \mathcal{N}, k \in \mathcal{K} \quad (4.2)$$

$$\sum_{k \in \mathcal{K}} x_{ij}^k \leq u_{ij} y_{ij} \quad \forall (i,j) \in \mathcal{A} \quad (4.3)$$

$$x_{ij}^k \leq b_{ij}^k y_{ij} \quad \forall (i,j) \in \mathcal{A}, k \in \mathcal{K} \quad (4.4)$$

$$y_{ij} \in \{0, 1\} \quad \forall (i,j) \in \mathcal{A} \quad (4.5)$$

$$x_{ij}^k \in \mathbb{R}_+ \quad \forall (i,j) \in \mathcal{A}, k \in \mathcal{K} \quad (4.6)$$

The objective (4.1) is to minimize the total cost of servicing the demand divided into the fixed costs of opening arcs for service and the variable cost of transport demand units on the open arcs. Constraints (4.2) are the traditional flow balance requirements imposed on the commodities and constraints (4.3) and (4.4) are referred to as the weak and strong forcing constraints respectively, imposing capacity restrictions on the flow of commodities on the individual arcs. Although the strong forcing constraints are redundant in an integer feasible solution, they can significantly improve the strength of the LP relaxation bound leading to better performance in the branch-and-bound search as described below. Finally, constraints (4.5) and (4.6) restrict variable domains.

The design constraints of the MCND model may be generalized by imposing the additional constraint $y \in Y$ with y being the vectorized form of the y_{ij} 's and Y being a set defining restrictions on the topology of the network, e.g., degree constraints on nodes. An example of a definition of Y restricting the topology of the network to consist of a series of cycles is provided in section 4.2.

Observe that for a feasible design vector y , the MCND reduces to a capacitated multicommodity network flow problem (see e.g., Ahuja et al. (1993) for a comprehensive treatment). This problem is at the core of network design problems and indeed arises as a sub-problem in many solution approaches based on decomposition and relaxation as well as being used in many practical applications within transportation (see e.g., Hane et al. (1995) for an example in aircraft fleet assignment) and telecommunication. A variation of the problem requiring indivisibility of commodities is also sometimes seen. Barnhart et al. (2000) discuss this problem

in further detail and also present a branch-and-price-and-cut algorithm to solve a series of bandwidth packing problems.

4.1.2 Applications and Solution Approaches

Applications for which model (4.1)–(4.6) has been employed are both plentiful and diverse as highlighted by e.g., Balakrishnan et al. (1997) and Ahuja et al. (1995). These include telecommunication network design, circuit board design, distribution planning, capital investment planning and of course freight transport network design. Gendron et al. (1999), Minoux (1989), Magnanti and Wong (1984) provide three additional surveys where the latter emphasizes applications in transportation planning.

One of the main challenges associated with solving a large class of network design problems using standard polyhedral methods (such as linear programming) is the fact that they have very poor integer properties which is one of the primary reasons that solving them using standard branch-and-bound is inefficient. The term *integer properties* relates to the gap between the lower bound obtained from an LP relaxation of a formulation and the value of an integer optimal solution. Several authors have noted this issue including Gendron and Crainic (1994), Crainic et al. (2001) and the gaps have been quantified as 20% on average for the weak formulation solving typical test instances (e.g. Gendron, 2001). For the LP relaxation of the strong formulations (or equivalent relaxations such as Lagrangian based ones) the gap is within 9% which is still quite large. The result is that a lot of time is spent closing the gap when searching the branch-and-bound tree and that only limited pruning is achieved leading to very large B&B trees. Some authors (e.g., Holmberg and Yuan, 1998, 2000, Crainic et al., 2001) have approached this problem by developing heuristics and employing alternative relaxation schemes such as Lagrangian based relaxation which has proved successful for the MCND problem (Gendron and Crainic, 1994). Other authors (Crainic and Gendreau, 2002, Gendron et al., 2003) seek to entirely avoid the problem by using (meta)heuristics.

The main reason for the poor integer properties is the fact that supply and demand for capacity rarely match. As there is no incentive to pay for additional (unused) capacity, this leads to LP relaxations in which a solution will match demand for capacity at the lowest possible level leading to fractional values for a large number of discrete-valued design variables. This issue is one of the motivating factors for the pursuit of alternative formulations which seek to avoid the problems associated with poor relaxations through reformulations and decision aggregation.

Both these techniques are central to the development of the presentation in this chapter.

4.2 Classical Network Design Extended

A major limitation of the MCND model (4.1)–(4.6) is the lack of asset balancing. Solutions to the MCND tend to exhibit a tree-like structure essentially requiring significant repositioning (empty journeys) to ensure asset balancing and offer continued service. Obviously, this is not a desirable property of a service network that needs to operate around the clock at the lowest possible cost and does not allow time for repeated asset repositioning.

A simple extension of the MCND model is the introduction of design balance constraints represented by constraint set (4.11) in the model ((4.7)–(4.13) below. These constraints impose requirements to cyclic routes which results in a network that essentially consists of a series of simple cycles, each of which can be thought of as being operated by a single vessel in a continually respecting loop. It is assumed that all vessels have equal characteristics. The extended model is referred to as the balanced multicommodity capacitated network design (BMCND) model.

$$(\text{BMCND}) \quad \min \sum_{(i,j) \in \mathcal{A}} f_{ij} y_{ij} + \sum_{k \in \mathcal{K}} \sum_{(i,j) \in \mathcal{A}} c_{ij}^k x_{ij}^k \quad (4.7)$$

s.t.

$$\sum_{j: (i,j) \in \mathcal{A}} x_{ij}^k - \sum_{j: (j,i) \in \mathcal{A}} x_{ji}^k = d_i^k \quad \forall i \in \mathcal{N}, k \in \mathcal{K} \quad (4.8)$$

$$\sum_{k \in \mathcal{K}} x_{ij}^k \leq u_{ij} y_{ij} \quad \forall (i,j) \in \mathcal{A} \quad (4.9)$$

$$x_{ij}^k \leq b_{ij}^k y_{ij} \quad \forall (i,j) \in \mathcal{A}, k \in \mathcal{K} \quad (4.10)$$

$$\sum_{j: (i,j) \in \mathcal{A}} y_{ij} - \sum_{j: (j,i) \in \mathcal{A}} y_{ji} = 0 \quad \forall i \in \mathcal{N} \quad (4.11)$$

$$y_{ij} \in \{0, 1\} \quad \forall (i,j) \in \mathcal{A} \quad (4.12)$$

$$x_{ij}^k \in \mathbb{R}_+ \quad \forall (i,j) \in \mathcal{A}, k \in \mathcal{K} \quad (4.13)$$

Model objective function and constraints remain identical to the corresponding constraints in MCND although with the addition of constraint set (4.11) enforcing design balance in all nodes of the graph. Thus, in reference to the generic version of the MCND model, the BMCND is a simple extension with $Y = \{y_{ij} | \sum_{j:(i,j) \in \mathcal{A}} y_{ij} - \sum_{j:(j,i) \in \mathcal{A}} y_{ji} = 0, i \in \mathcal{N}, (i,j) \in \mathcal{A}\}$.

Design balance has been utilized in several works addressing freight transportation problems. Pedersen et al. (2009) discuss the problem in a the general context of service network design emphasizing applications in maritime and air transport. The authors propose a Tabu search heuristic to solve the problem and report results for instances with up to 400 commodities in graphs with 30 nodes and 700 arcs. An application within vehicle routing for package delivery is presented by Smilowitz et al. (2003) who propose several LP relaxation based rounding heuristics to solve the problem. Other works impose design balance requirements in decomposed models at the master and/or sub-problem level including Hane et al. (1995) for the aircraft fleet assignment problem and Barnhart and Schneur (1996), Kim et al. (1999) for an express package SND problem.

One of the challenges of introducing design balance requirements again arises in connection with the use of the LP relaxation in a branch-and-bound setting to solve the problem. The fractionalities that are a problem in the MCND are propagated by the balance requirements thus compounding the issue of poor lower bounds and excessive time spent building the branch-and-bound tree due to a lack of pruning options.

A Lagrangian relaxation based approach to solving model (4.7)–(4.13) is discussed in Appendix A. In the proposed approach, constraints (4.9) and (4.10) are relaxed in a Lagrangian fashion to obtain a series of independent sub-problems. This leads to a relaxed problem that is feasible with respect to integer requirements, but may not satisfy all capacity constraints which is to be contrasted with the LP relaxation which satisfies all constraints except the integer requirements. The Lagrangian dual problem is solved using a stabilized cutting plane method. Although the Lagrangian relaxation has the same theoretical bound as the LP relaxation, it initially converges faster than traditional dual simplex and provides valuable primal and dual information. The exploitation of this information in a heuristic is discussed together with the integration into a parallel branch-and-bound framework.

4.2.1 A Cycle Based Reformulation

The variables of the BMCND model may be redefined such that freight flows are modeled as complete origin-destination paths rather than individual link flows thus eliminating the flow balance constraints (4.8). Similarly, design variables can be recast to represent vessel routes which, due to the design balancing requirements (4.11), correspond to closed cycles (rotations in the terminology of liner shipping). This alternative definition of the design variables emphasizes the link to physical assets in the form of vessels operating rotations and is thus very appealing from the perspective of modeling the liner SND.

Let $a_{ijr} = 1$ if arc $(i, j) \in \mathcal{A}$ is serviced in route $r \in \mathcal{R}$ with route r defined as a sequence of arcs $(i_1, i_2), (i_2, i_3), \dots, (i_n, i_1)$ and $a_{ijr} = 0$ otherwise. Furthermore, let $f_r = \sum_{(i,j) \in r} f_{ij}$ represent the cost of operating route $r \in \mathcal{R}$. Maintaining the flow variables as defined in the original BMCND and letting $y_r = 1$ if a cyclic route $r \in \mathcal{R}$ is operated and $y_r = 0$ otherwise, the cyclic equivalent formulation of BMCND can be stated as follows:

$$(\text{cBMCND}) \min \sum_{r \in \mathcal{R}} f_r y_r + \sum_{k \in \mathcal{K}} \sum_{(i,j) \in \mathcal{A}} c_{ij}^k x_{ij}^k \quad (4.14)$$

s.t.

$$\sum_{j: (i,j) \in \mathcal{A}} x_{ij}^k - \sum_{j: (j,i) \in \mathcal{A}} x_{ji}^k = d_i^k \quad \forall i \in \mathcal{N}, k \in \mathcal{K} \quad (4.15)$$

$$\sum_{k \in \mathcal{K}} x_{ij}^k \leq u_{ij} \sum_{r \in \mathcal{R}} a_{ijr} y_r \quad \forall (i, j) \in \mathcal{A} \quad (4.16)$$

$$x_{ij}^k \leq b_{ij}^k \sum_{r \in \mathcal{R}} a_{ijr} y_r \quad \forall (i, j) \in \mathcal{A}, k \in \mathcal{K} \quad (4.17)$$

$$\sum_{r \in \mathcal{R}} a_{ijr} y_r \leq 1 \quad \forall (i, j) \in \mathcal{A} \quad (4.18)$$

$$y_r \in \{0, 1\} \quad \forall r \in \mathcal{R} \quad (4.19)$$

$$x_{ij}^k \in \mathbb{R}_+ \quad \forall (i, j) \in \mathcal{A}, k \in \mathcal{K} \quad (4.20)$$

Constraints (4.15)–(4.17) are similar to the corresponding constraints in BMCND with the exception that occurrences of y_{ij} are now replaced with $\sum_{r \in \mathcal{R}} a_{ijr} y_r$. An additional constraint set (4.18) has been added representing route packing constraints to enforce that any arc (i, j) can

only be included in one active operated route. The reformulation of the design variables eliminates the need for explicitly representing the design balance constraints (4.11) in the model. However, it should be clear that the reduced complexity in terms of fewer constraints for the cBMCND model comes at a price of many more variables as the number of potential routes \mathcal{R} becomes extremely large. For realistically sized problems it is intractable to explicitly enumerate all feasible routes and techniques such as delayed column generation (see e.g., Barnhart et al., 1998) are required.

Assuming the graph on which the routes are defined has a time-space structure, the cBMCND model (and the equivalent BMCND model) has extended the original MCND model such that it now captures the following requirements (* indicates new dimensions compared to the previous model):

- All demand must be met
- Demand varies over the planning horizon
- Servicing demand must respect asset capacities
- Routes must be cyclic *
- Port time windows are respected *

It is worth noting that in order for the BMCND to capture port time windows, the underlying graph G must have a time-space structure and an appropriate time discretization must be chosen depending on the size of port time windows.

Strictly speaking, as there is no representation of individual vessels, transshipment of freight is not really well-defined. However, it is intuitive in the cBMCND to adopt a view of freight being transshipped between routes and as such it could be argued that cBMCND (and the equivalent BMCND) formulation models the possibility of transshipping freight. It is important to remember, however, that freight transshipment is very expensive relative to the cost of transportation which limits the usefulness of the model somewhat as it fails to fully capture the economical impact. The next section seeks to address this issue.

4.2.2 Adding Layers of Complexity

The preliminary experience on solving the BMCND using exact methods reported in Appendix A suggested that pursuing exact approaches to balanced network design formulations requires additional work to find use in real-world applications. However, for the sake of completeness and to support the transition to the later models, a multi-layer version

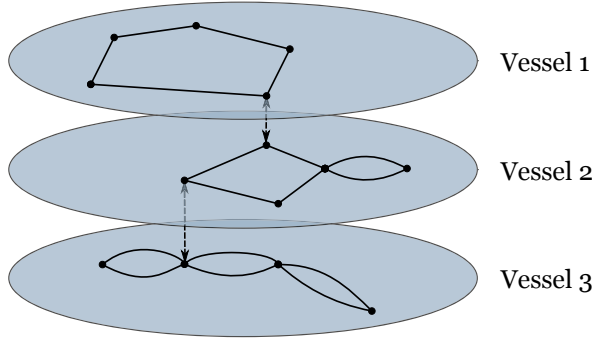


Figure 4.1: Illustration of the layered network interpretation of the BM-CND with multiple distinct vessels and freight transshipment (vertical lines).

of the BMCND is included below. Essentially, the multi-layer BMCND (MLBMCND) model is obtained by extending the BMCND to capture the movement of individual vessels as well as freight flows on these vessels. The term multi-layer arises from the interpretation of this new model as being multiple network layers, one for each vessel, linked together by the freight flows and transshipment of this freight among vessels at different layers in the network. This interpretation has been visualized in Figure 4.1 where each layer represents a route of an individual vessel based on an underlying graph (not included to improve readability). Dashed vertical lines represent transshipment arcs (bidirectional) where two routes visit a common port and freight can be transshipped from one route to another.

In terms of modeling the liner SND, the multi-layer model enables the capture of the following requirements and constraints (* indicates new dimensions compared to the previous model):

- All demand must be met
- Demand varies over the planning horizon
- Servicing demand must respect asset capacities
- Routes must be cyclic
- There can be multiple intersecting routes *
- Vessel-port compatibility is respected *
- Cargo can be transshipped between vessels and the cost of this is captured. *
- Port time windows are respected

Before presenting the model, a few additional definitions are necessary.

Let \mathcal{V} denote a set of vessels and let y_{ijv} , $(i, j) \in \mathcal{A}_v$ correspond to the design variables defined in section 4.1 but now specified for each individual vessel $v \in \mathcal{V}$ on the associated graph $G_v = (\mathcal{N}_v, \mathcal{A}_v)$. The set of all nodes is denoted $\mathcal{N} = \cup_{v \in \mathcal{V}} \mathcal{N}_v$. Similarly, let x_{ijv}^k denote the amount of commodity $k \in \mathcal{K}$ flowing on arc $(i, j) \in \mathcal{A}_v$ on vessel $v \in \mathcal{V}$ with a cost of c_{ijv}^k incurred per unit of flow. Furthermore, let s_{iv}^k denote the amount of commodity $k \in \mathcal{K}$ that is discharged from vessel $v \in \mathcal{V}$ at node $i \in \mathcal{N}_v$. A cost c_{iv}^k is associated with the discharge of a commodity $k \in \mathcal{K}$ and the cost may depend on both vessel v and node i . With an appropriate definition of these discharge costs, it is possible to capture the costs associated with transshipment of individual commodities. Using these definitions, the MLBMCND model can be expressed as follows:

(MLBMCND)

$$\min \sum_{v \in \mathcal{V}, (i,j) \in \mathcal{A}_v} f_{ijv} y_{ijv} + \sum_{v \in \mathcal{V}, k \in \mathcal{K}} \sum_{(i,j) \in \mathcal{A}_v} c_{ijv}^k x_{ijv}^k + \sum_{v \in \mathcal{V}, i \in \mathcal{N}_v, k \in \mathcal{K}} c_{iv}^k s_{iv}^k \quad (4.21)$$

s.t.

$$\sum_{v \in \mathcal{V}, j: (i,j) \in \mathcal{A}_v} x_{ijv}^k - \sum_{v \in \mathcal{V}, j: (j,i) \in \mathcal{A}_v} x_{jiv}^k = d_i^k \quad \forall i \in \mathcal{N}, k \in \mathcal{K} \quad (4.22)$$

$$\sum_{j: (j,i) \in \mathcal{A}_v} x_{jiv}^k - \sum_{j: (i,j) \in \mathcal{A}_v} x_{ijv}^k \leq s_{iv}^k \quad \forall v \in \mathcal{V}, k \in \mathcal{K}, i \in \mathcal{N}_v \quad (4.23)$$

$$\sum_{k \in \mathcal{K}} x_{ij}^k \leq u_{ij} y_{ijv} \quad \forall v \in \mathcal{V}, (i,j) \in \mathcal{A}_v \quad (4.24)$$

$$x_{ijv}^k \leq b_{ij}^k y_{ijv}^k \quad \forall v \in \mathcal{V}, (i,j) \in \mathcal{A}_v, k \in \mathcal{K} \quad (4.25)$$

$$\sum_{j: (i,j) \in \mathcal{A}_v} y_{ijv} - \sum_{j: (j,i) \in \mathcal{A}_v} y_{jiv} = 0 \quad \forall i \in \mathcal{N}_v, v \in \mathcal{V} \quad (4.26)$$

$$\sum_{(i,j) \in \mathcal{A}_v} \delta_{ijv} y_{ijv} \leq 1 \quad \forall v \in \mathcal{V} \quad (4.27)$$

$$y_{ijv} \in \{0, 1\} \quad \forall (i,j) \in \mathcal{A}_v, v \in \mathcal{V} \quad (4.28)$$

$$x_{ijv}^k \in \mathbb{R}_+ \quad \forall (i,j) \in \mathcal{A}_v, k \in \mathcal{K}, v \in \mathcal{V} \quad (4.29)$$

$$s_{iv}^k \in \mathbb{R}_+ \quad \forall v \in \mathcal{V}, i \in \mathcal{N}_v, k \in \mathcal{K} \quad (4.30)$$

The objective function (4.21) of the MLBMCND is very similar to the objective of the BMCND with the exception that design costs f_{ijv} , $v \in \mathcal{V}$, $(i,j) \in \mathcal{A}_v$ are now divided onto individual vessels and a new term representing the cost of transshipping commodities has been included. With the exception of constraint sets (4.27) and (4.23), the remaining constraints are simply the balancing, and forcing constraints of the BMCND expressed for individual vessels. Constraint set (4.23) has been included

to keep track of the amount of freight that is discharged from individual vessels at each node of the network.

What remains is to ensure that each vessel performs a route that does not contain any sub-tours. Assuming a time-space structure in the underlying graph $G = (\mathcal{N}_v, \mathcal{A}_v)$ for each vessel $v \in \mathcal{V}$ it is possible to replace the traditional exponential set of sub-tour elimination constraints with a single count-line constraint for each vessel. The count-line constraints expressed in (4.27) simply count the number of times a particular vessel crosses a specific point in time indicated by $\delta_{ijv} = 1$ if arc $(i, j) \in \mathcal{A}_v$ crosses this point and $\delta_{ijv} = 0$ otherwise. The sum of all crossing counts (arcs crossing the count line) is then restricted to one thereby enforcing that each vessel must operate a closed cycle and restricting the length of this cycle to the duration of the planning horizon. A simple extension of these constraints realized by including one additional vessel specific binary variable on the right hand side and enforcing equality will allow for the capture of initial costs (e.g., time charter) incurred by the deployment of a particular vessel.

Andersen et al. (2009) conducted an analysis of a variation of MLBM-CND model by comparing the strength of a series of compact and extended formulations based on various forms of decomposition. The four formulations investigated were based on modeling 1) arc-based flow and design, 2) path-based flow and arc-based design, 3) arc-based flow and cycle-based design, and 4) path-based flow and cycle based design. All the extended formulations (path- or cycle-based) were solved using complete variable enumeration. The analysis concluded that the cycle-based formulations provided stronger LP relaxations and thus also required less effort to solve. Furthermore, comparisons with simplified versions of the model excluding among others the design balance constraints show significant reductions (more than 90%) in the required solution time. This again supports the note in section 4.2 which suggested that design balance requirements tend to propagate fractionalities and thus complicate model solution using standard polyhedral methods.

4.3 Decomposing using Specific Problem Properties

Up until now, the discussion has focused primarily on compact formulations for a service network design problem with only two simple components in the form of demand flow and route design and two corresponding constraint types relating to these plus an additional type tying

these together. Modeling approaches have been based on a more or less straight-forward mapping of the components of the service network by observing that vessels operate routes consisting of journey legs with commodities flowing on these legs. However, as evidenced by the experiments with these compact formulations, alternative approaches are necessary if the liner SND is to be successfully modeled and solved in real-world cases. Also, the straight-forward approaches fail to capture some of the structural properties that may exist in the problem and which may be exploited to develop more efficient algorithms.

Although the multi-layered model is impractical from the perspective of using standard polyhedral methods and even solving the more simple single-layer version (BMCND) using relaxation is challenging, the structure of the models suggest a slightly different approach based on layer decomposition. Rather than simultaneously modeling route design and general freight flows, two simplifying assumptions will allow for a model in which routes and aggregated freight flows are represented by a single variable each. Assuming that transshipments are not allowed to impact the design of the service network and thus, that demand forecast implicitly specifies requirements for direct services, it is possible to substitute general freight flows with vessel specific freight allocations. Further assuming that vessels generally become empty at the major hub ports of the network means that rather than enforcing that a single route must be a closed cycle with a length corresponding to the length of the planning horizon, routes can now be represented by a series of shorter sub-routes each of which start and end at a hub port. Using problem information it is possible to derive additional constraints on feasible and practical route lengths based on freight volume since it is generally undesirable to visit a large number of ports on a long route if only small amounts of freight are serviced (picked up or delivered) at each port. Service level requirements may help to further support the case for shorter routes since it is considered inefficient from a competitive standpoint to offer services with too long transit times. As the the number of possible routes increases exponentially with the length of these routes, shorter routes can significantly improve tractability of the modeling approach. It is noted that the assumption about vessels becoming empty need not necessarily be constraining as the model is general in terms of representing routes and thus allows for routes spanning the whole planning horizon. Such routes will allow arbitrary demand flows within the boundaries of a single route. For a further discussion on the modeling approach, the reader is referred to chapter 5.

Before presenting the model, a few brief definitions will be required as the concepts of the model are significantly different from those of the

models in the previous sections.

Similar to the cBMCND model, let y_r define the decision to operate a route $r \in \mathcal{R}$ such that $y_r = 1$ if route r is operated and $y_r = 0$ otherwise. A route $r \in \mathcal{R}$ need not form a simple cycle as for the cBMCND, in fact it need not even form a closed cycle but it must begin and end at a hub node $i \in \mathcal{N}_H$ of the so called hub-graph $\mathcal{G}_H = (\mathcal{N}_H, \mathcal{A}_H)$. The set of all routes can be partitioned per vessel class such that $\mathcal{R} = \cup_{v \in \mathcal{V}} \mathcal{R}_v$ with \mathcal{V} representing a set of vessel classes each with a number of available vessels n_v , $v \in \mathcal{V}$. With each route there is an associated cost c_r incurred by operating route $r \in \mathcal{R}$.

The decision to service a commodity group is represented by x_{gr} , $g \in \mathcal{G}_r$, $r \in \mathcal{R}$ with \mathcal{G}_r representing the set of commodity groups that are feasible for route r . Also, let w_{vi+} (w_{vi-}) represent the number of vessels $v \in \mathcal{V}$ occupying a waiting arc out of (respectively into) node $i \in \mathcal{N}_H$ which can be interpreted as the number of vessels staying idle at a port. A cost c_{vi} is associated with a vessel $v \in \mathcal{V}$ waiting at a hub port $i \in \mathcal{N}_H$. Furthermore, define by $\mathcal{R}_{v(i)+} \subseteq \mathcal{R}_v$ the subset of routes of vessel class v starting at hub node $i \in \mathcal{N}_H$ (similarly $\mathcal{R}_{v(i)-}$ are the routes ending at hub node i).

The model, dubbed the Liner Service Network Design (LSND) model, is in its entirety defined as follows:

$$(\text{LSND}) \quad \min \sum_{r \in \mathcal{R}} c_r y_r + \sum_{v \in \mathcal{V}} \sum_{i \in \mathcal{N}_H} c_{vi} w_{vi+} \quad (4.31)$$

s.t.

$$\sum_{r \in \mathcal{R}} \sum_{g \in \mathcal{G}^k} x_{gr} = 1 \quad \forall k \in \mathcal{K} \quad (4.32)$$

$$\sum_{g \in \mathcal{G}_r} x_{gr} \leq y_r, \quad \forall r \in \mathcal{R} \quad (4.33)$$

$$\sum_{r \in \mathcal{R}_v(i)^+} y_r + w_{vi+} - \sum_{r \in \mathcal{R}_v(i)^-} y_r - w_{vi-} = 0 \quad \forall v \in \mathcal{V}, i \in \mathcal{N}_H \quad (4.34)$$

$$\sum_{r \in \mathcal{R}_v} \delta_r y_r + \sum_{i: w_{vi+} \in \delta(\mathcal{W}_v)} w_{vi+} \leq n_v \quad \forall v \in \mathcal{V} \quad (4.35)$$

$$x_{gr} \in [0, 1], \quad \forall r \in \mathcal{R}, g \in \mathcal{G}_r \quad (4.36)$$

$$y_r \in \{0, 1\} \quad \forall r \in \mathcal{R} \quad (4.37)$$

$$w_{vi+} \in \mathbb{Z}_+ \quad v \in \mathcal{V}, \forall i \in \mathcal{N}_H. \quad (4.38)$$

The objective (4.31) of LSND is to minimize the total cost of operating the set of selected routes and cost of having vessels stay idle in ports. The covering constraints (4.32) enforce that all commodities must be serviced in the selected commodity groups and (4.33) ensure that a corresponding set of routes is selected to provide sufficient capacity to meet the demand. Constraints (4.34) are balance constraints for the assets expressed in each node of the hub graph for each vessel class. Finally, constraint set (4.35) are the count line constraints enforcing that the number of vessels within each vessel class required to operate the selected routes cannot exceed the number of available vessels.

It is clear that there are challenges associated with the determination of feasible routes \mathcal{R} and the sets of associated commodity groups \mathcal{G}_r , $r \in \mathcal{R}$. Furthermore, from the perspective of using a delayed column generation approach, there is no direct dual information available to price out new route variables and the corresponding commodity group sub-problems depend on the individual route on which it will be serviced. However, routes can be priced out using dual approximation in either a dynamic programming or heuristic setting. Similarly, determining new commodity groups requires the solution of a sub-problem with a multidimensional knapsack structure. These problems are, however, very small. The liner

SND model and the solution approach is described in further detail in the included Paper 1 in chapter 5. This chapter also contains results from computational experiments based on the application of LSND to a series of adapted real-world cases.

In terms of modeling the liner SND the the LSND model captures the following aspects:

- All demand must be met
- Demand varies over the planning horizon
- Servicing demand must respect asset capacities
- Routes must be cyclic
- There can be multiple intersecting routes
- Vessel-port compatibility is respected
- Port time windows are respected

It is observed that transshipments are no longer handled with reference to the above discussion regarding the undesirability of these in the design of the service network of the concrete application. Although not explicitly included in the implementation proposed in Paper 1, the LSND model can actually capture a few additional aspects of the liner SND. These include

1. Multiple *capacity* dimensions on vessels
2. Actual capacity may depend on route
3. Costs depend on vessel cruising speed
4. Demand may depend on service level

Furthermore, the model in (4.31)–(4.38) can be extended to capture some degree of service level requirements for the individual ports. In contrast to many of the above aspects which must be enforced/captured at the route generation sub-problem level, port service level requirements must be included in the master problem. Given these possible extensions, the LSND is actually quite general. Add to this the fact that many of the above extensions incur only limited additional computational complexity in the sub-problems and the LSND becomes a realistic proposal for a model of the liner SND problem.

4.3.1 The Concept of Composites

The decomposition approach adopted in the LSND shares ideas with the composite variable paradigm. Rather than looking at variables as being abstractions of decisions relating to a single resource (such as the opening of a service in models MCND and (ML)BMCND), composite

variable modeling adopts a more holistic view by defining variables covering overall objectives. Thus, the central idea behind composite variable modeling is that the selection of a single composite achieves an overall objective. In this regard, the concept of a commodity group can be considered a composite as it ensures the complete service of the individual commodities part of that group. Other examples of the use of composite variables and a discussion about the impact of this approach can be found in Armacost (2000), Armacost et al. (2002). Here the idea of composite variables is developed for an express package SND in which a composite represents an aircraft route to which capacity is assigned in the form of multiple aircraft ensuring the coverage of all demand in the visited airports. The use of composites in this work is facilitated by a very special structure making it possible to a priori construct all relevant composite variables. Another example is the use of composite variable modeling in connection with service part logistics presented by Cohn and Barnhart (2006). In this case, the selection of a composite corresponds to the simultaneous satisfaction of all demand for a particular service part in a distribution network based on warehouses and service part consumers. Regardless of the application however, it is crucial that the choice of a composite variable allows for efficient enumeration of these, either explicitly a priori or implicitly in a branch and price setting. With the correct definition of composite variables, these can significantly improve the tractability of the models using traditional polyhedral methods as symmetries are reduced and formulations are stronger with respect to e.g., linear programming relaxations.

4.4 Models with a Set Partitioning Structure

Following the principle of decision aggregation one step further, but using a slightly different decompositional approach, the result becomes a model in which a single variable determines both freight allocations and the associated routing for an single vessel. Actually, the routing is derived based on the sequence of pickups and deliveries of commodities, with the construction of the sequence taking into account feasibility of the route.

Let $\mathcal{R} = \cup_{v \in \mathcal{V}} \mathcal{R}_v$ denote a set of all candidate routes for a fleet of vessels \mathcal{V} with feasible routes \mathcal{R}_v specified for each vessel $v \in \mathcal{V}$. Furthermore, \mathcal{K} define the set of commodities to be serviced as in the previous models and let $\mathcal{R}^k \subset \mathcal{R}$ be the set of routes that service commodity $k \in \mathcal{K}$. Finally, define by $y_r = 1$ the decision to operate a route $r \in \mathcal{R}$ at cost

c_r and $y_r = 0$ otherwise. With these definitions the set partitioning formulation of the SND can be expressed as follows:

$$(\text{SP}) \ z = \min \sum_{r \in \mathcal{R}} c_r y_r \quad (4.39)$$

s.t.

$$\sum_{r \in \mathcal{R}_v} y_r = 1 \quad \forall v \in \mathcal{V} \quad (4.40)$$

$$\sum_{r \in \mathcal{R}^k} y_r = 1 \quad \forall k \in \mathcal{K} \quad (4.41)$$

$$y_r \in \{0, 1\} \quad \forall r \in \mathcal{R} \quad (4.42)$$

$$(4.43)$$

The objective (4.39) is to minimize the cost of operating the selected routes. An alternative to this definition of the objective is a profit based perspective although this would violate the requirement about ensuring consistency in input and decisions stated in the introduction to this chapter. Constraint set (4.40) ensures that only one route is selected for each vessel and constraint set (4.41) enforces that all commodities should be covered by exactly one route.

Set partitioning adopts a strict "commodity" view by assuming all relevant information is contained in commodity time windows. However, service level is only implicitly captured and route inter-dependence (e.g., synchronization to facilitate transshipment of cargo) is ignored but may to some extent be captured by extending the model as described below. General forms of route inter-dependency, however, adds significant complexity in sub-routines for route building as it introduces additional constraints. Thus, only if pre-defined patterns can be specified, will this approach work and allow modeling some degree of route inter-dependency.

The simple set partitioning model above may be extended by a number of additional constraints governing the interaction between routes of the different vessels. An example of such constraints is the use of visit patterns introduced by Sigurd et al. (2005) to capture service level requirements. These constraints actually introduce route inter-dependence but at the master problem level rather than in the sub-problems.

With the visit pattern extension, the SP model can capture the following aspects of the liner SND:

- All demand must be met
- Demand varies over the planning horizon

- Servicing demand must respect asset capacities
- Multiple capacity dimensions on vessels
- Routes must be cyclic
- There can be multiple intersecting routes
- Vessel-port compatibility is respected
- Port time windows are respected
- Service level requirements are imposed on ports

4.4.1 Solution Approaches

The simple formulation of the SP model (4.39)–(4.42) is achieved by delegating a significant amount of the complexity to a sub-problem responsible for the generation of feasible route sets \mathcal{R}_v , $v \in \mathcal{V}$. Although SP may contain much fewer constraints than the previous compact formulations, the model typically contains a very large number of variables. To understand the challenges that are faced in the sub-problem it is sufficient to realize that e.g., the number of potential route candidates \mathcal{R}_v for single vessel is exponential in the number of commodities $|\mathcal{K}|$. Even for moderately size problems, it is not feasible to explicitly enumerate all these route candidates, nor is it desirable as many will never be considered selected in good solutions. The solution to this problem is again to adopt delayed column generation, where the complexity of determining new attractive route candidates is delegated to a series of sub-problems.

Different approaches to solving the route generation sub-problems can be adopted and they fall into two categories; exact and approximate. Exact approaches include dynamic programming or the use of linear or integer programs (as was the case for the LSND) to determine attractive route candidates. For simple sub-problems, generating new route candidate can reduce to solving a series of shortest path problems over appropriately constructed networks. Typically, however, sub-problems are more complex and often involve the solution of resource constrained shortest path problems (see e.g., Irnich and Villeneuve (2006), Boland et al. (2006)) that are much more computationally demanding (versions requiring elementarity of the path are actually NP-hard in general, see Garey and Johnson (1979) for a treatment of complexity theory). Approximate approaches typically employ some heuristic procedure to determine attractive routes and may be combined with exact approaches that are only employed when the heuristic fails to produce new route candidates. Some of the most successful approaches to solving e.g., vehicle routing problems are based on the use of delayed column generation with sub-problems combining heuristics and exact methods.

4.5 Summary

Although the models discussed above are both diverse and rich, none of them captured all of the requirements to a liner service network design set forward at the beginning of the chapter. While the LSND model provided the potential to capture many of the requirements, it still failed to explicitly include the possibility of freight transshipment. As it turns out, transshipments links routes through individual commodities in a way that makes it difficult to employ any of the traditional modeling and decomposition techniques utilized above. It was shown in the MLBMCND model that it is possible to capture transshipments and the associated costs using extended formulations and straight forward variable definitions. However, it was also clear that such an approach is not computationally tractable and thus of limited practical value at this time.

It is relevant to question whether the presence of transshipments is even a desirable feature in a model, given the high costs incurred by these additional handling operations and of course the resulting model complexity. It is not unrealistic to imagine that even if transshipments are initially ignored, resulting service network designs may still facilitate a sufficient set of transshipment options. However, to fully evaluate the consequence of ignoring the possibility of freight transshipment, it is necessary first to be able to model them. The challenge of course is that the conclusion is likely to depend on the type of service network that is operated.

Given the experience from the above presentation of a series of very different approaches to modeling and solving the liner service network design, it currently seems that heuristics or possibly hybrid methods combining exact and heuristic techniques is the most viable path forward. Also, a new way of thinking of problem decomposition may be required to fully capture the rich requirements imposed on the liner service network design. Initial steps were provided in the LSND model using a route first, allocate (freight) second principle based on the observation that when routes are known, much more complex aspects can be handled in the allocation of freight to these routes.

CHAPTER 5

Paper 1: Service Network Design in a Liner Container Feeder Application

Martin W. Andersen,
Department of Transport,
Technical University of Denmark

Cynthia Barnhart,
Civil and Environmental
Engineering and Engineering
Systems,
Massachusetts Institute of
Technology

Oli B.G. Madsen,
Department of Transport,
Technical University of Denmark

Abstract

Liner based container shipping service is a key component of many modern supply chains and as such, carriers are constantly facing demands for optimizing the services they provide. This paper considers the problem of designing the service network and schedule for a liner container feeder operator with the objective of minimizing the cost of maintaining the resulting network. We propose an algorithm that exploits some of the structural properties of the problem to decompose the model into a series of sub-problems. The main idea builds on a route pool which is iteratively augmented using a dual based heuristic and a master problem which is dynamically expanded as new routes and freight allocations are generated. Computational experiments based on a real-world case show the feasibility of our algorithm with respect to solving realistically sized problems. Results indicate solutions with high utilization of vessel capacity can be obtained within a reasonable time.

5.1 Introduction

Maritime freight transport has been referred to as the backbone of international trade and a key player in the development of the modern supply chains. Freight is carried in one of three major modes; bulk, liquid and containerized of which the latter represents high value-low volume commodities. In recent years, transportation of containerized goods has seen very large annual growth rates (UNCTAD, 2008) and containerization continues to play an important role in the development and expansion of globalized supply chains with specific emphasis on manufactured goods, (Broeze, 2002, Levinson, 2006). With increasing freight volumes and resulting increases in vessel fleets, logistic service providers face new challenges in both strategic, tactical as well as operational planning of their service network. To remain competitive in a market with constantly diminishing freight rates, logistic service providers must maintain a high level of responsiveness to changes in demand which in turn requires the ability to efficiently update and evaluate their service network. This need is compounded by recent trends toward requirements for a higher level of specialization in product offerings and a continued need to increase asset utilization to counter the falling freight rates as pointed out by Hingorani et al. (2005). Finally, customers are increasingly demanding high reliability and integrated services leading to more complex service networks for the logistics service providers.

Where deep-sea carriers are responsible for the main haul of most international transport of containerized goods measured in tonne-miles, feeder services typically carry out the first and last part of the sea-based journey in regions with lower levels of freight consolidation. Despite sharing many features with deep-sea carriers, feeder services often operate smaller vessels over shorter distances and under different asset management constraints. This offers a larger degree of flexibility in the schedule creation as well as the fleet deployment plans. The problem of designing the service network in a liner based container feeder operation will be the focus of this paper.

Two key planning problems are faced by container feeder service providers; 1) tactical service network design including fleet size and mix, collectively referred to as the *master schedule* problem and 2) definition of the operational schedule and the execution of this schedule. Thus, in this context, the operational schedule is a result of implementing the master schedule and running the daily business.

We shall limit our scope to the modeling and solving of the master schedule problem but, when relevant, notes will be given to aspects of the operational planning that might affect decisions regarding the master schedule. Development will be based on a concrete case and thus also exploit some of the structural properties present in this case. However, the overall model and solution approach is built on more general concepts and can thus be applied to a wider range of problems within both liner shipping as well as other applications of consolidation based service network design.

5.1.1 Previous Work

One of the earliest published studies of liner based container shipping was conducted by Weldon (1958) to evaluate the profitability of introducing containerization of general cargo. The study was based on the definition of a series of equipment and demand scenarios for which the total system cost was calculated. Later work by Weldon (1959) builds on the findings of Weldon (1958) by developing a simulation model which allows for the evaluation of different route and schedule definitions in an operational context based on changing demand conditions.

Despite early adoptions of operations research techniques, liner based maritime transport problems have received relatively little attention with research focused mainly on tramp and industrial shipping as evidenced by surveys by Ronen (1983), Ronen (1993) and Christiansen et al. (2004). Of the 136 references presented in these three surveys, 18 discuss liner

based freight transport problems. Similarities in problem characteristics, however, mean that algorithmic ideas from other modes of shipping as well as vehicle routing and pickup-and-delivery problems (see e.g. Toth and Vigo (2002)) also find application in service network design for liner container shipping.

In a two part study, Perakis and Jaramillo (1991) and Jaramillo and Perakis (1991) use a linear programming model to solve a fleet deployment problem with routes given a priori. Service frequency requirements are enforced through constraints on the minimum number of voyages performed on the serviced routes and deployments are derived from the continuous variables by rounding. The first paper Perakis and Jaramillo (1991) contains an extensive discussion of the cost structure for a liner based maritime network. Experimental results based on a real-world case show savings of 3% by operating with an optimized deployment plan. The work by Perakis and Jaramillo is extended by Powell and Perakis (1997) by solving the same model but with integer variables and the goal remains to minimize the total operating and lay-up costs for the fleet.

Sigurd et al. (2005) extend a basic set partitioning formulation to include explicit requirements on the lead time between pickup and delivery, recurrence, and separation of calls to a set of customer locations. This is achieved through the introduction of pre-defined feasible visit patterns. They adopt a delayed column generation approach and exploit the network structure in the studied case with a remotely located hub to reduce the complexity of their sub-problem. A practical case with 68 weekly commodities from 21 ports is solved over one and two-week planning horizons respectively showing savings 14.8% by allowing a longer planning period.

Using a variation of the classical capacitated multicommodity network design problem (see e.g. Ahuja et al. (1995)), Agarwal and Özlem (2008) model a ship scheduling and cargo routing problem and propose three heuristics to solve problems with up to 20 ports and three vessels classes. Weekly frequencies are enforced through the pre-allocation of a sufficient number of vessels during the definition of routes. In contrast to other papers discussed here, the model does allow for the transshipment of cargo between different vessels although without capturing the cost of these operations. Comparison of a greedy, a column generation, and a Benders decomposition based heuristic show that the performance of the latter two is comparable on small problems. On larger problems the Benders based heuristic performs best.

In this work we do not explicitly consider issues related to reverse lo-

gistics in the form of empty container repositioning. However, as also noted by e.g. Agarwal and Özlem (2008), it is possible to include empty flows as additional commodities thus competing for space with the profit bearing commodities. In connection with the reverse logistics issue, it is worth mentioning that empty container repositioning is typically a profit bearing activity in feeder networks as feeder carriers do not own containers themselves.

5.1.2 Contributions and Outline

In terms of modeling and solving the liner container feeder service network design problem the contributions of this paper include

- The development of an LP based heuristic for the feeder service network design problem capable of handling rich properties of the problem.
- Exploiting structural properties of the specific problem to mitigate some of the issues traditionally associated with employing standard polyhedral methods for network design, i.e., poor lower bound properties of the LP relaxation.
- Showing the tractability of our approach by solving a realistically sized real-world problem for a container feeder operator.

The remainder of this paper is organized as follows. Section 5.2 presents the tactical feeder network design problem in further detail and discusses the limitations and assumptions. A formal description of the feeder network design model and its decomposition will be provided in section 5.3 while section 5.4 will discuss the components of the decomposed problem further in the context of the solution approach adopted. This section will also describe the heuristic column generation procedure adopted. Computational results from solving a real-world case will be discussed in section 5.5 before concluding remarks and perspectives are given in section 5.6.

5.2 Liner Container Feeder Planning

The creation of a master schedule represents the integration of company strategies, competitive requirements as well as expectations to the future market. In essence, the feeder service provider must determine which locations to service, which routes to operate, and which vessel fleets to assign to the different routes. The decisions made will determine the

products that can subsequently be offered to the customers and, equally important, the level of service that can be provided. Several considerations govern the design of the master schedule but ultimately, the goal is to meet customer demand for freight transport with the highest level of service while satisfying corporate profit goals.

The master schedule is revised every three to six months in response to changes in demand patterns and volumes. Changes to the master schedule may include introducing both new routes as well as new vessels. In contrast to ocean liners who typically publish schedules that are fixed six months into the future, the ability to perform frequent and efficient updates to the schedule is a key competitive parameter for feeder service providers.

In the case considered in this paper, the feeder service provider is classified as a non-vessel owning common carrier meaning that the majority of the fleet is chartered on contracts ranging in term from anywhere between a few weeks to several years. The consequence of this type of operation is that it is possible for the feeder service provider to change the fleet size and composition within a relatively short period of time, thus providing a very high degree of flexibility in the master planning process. However, this also adds another dimension of complexity since charter rates are highly market dependent and may increase or decrease in expectation of future demand for container vessels. Although the contract negotiation aspect of the chartering is not considered in this work, the consequence of operating a chartered fleet is that vessels with similar characteristics may incur different costs.

Further facilitating the flexibility in designing the master schedule is the relative geographical closeness of the ports that are serviced in the feeder network. This results in higher frequency of calls in the individual ports and will typically mean that vessels can be redeployed to new routes within a short period of time.

5.2.1 Components of the Service Network

The two basic components of the service network is the set of ports that are served and the routes performed by vessels that provide the transport service.

In general, a physical *port* may consist of several terminals and the port itself can have associated time windows in which it is available for service to berthing vessels. Based on the role of a port, it will be classified as hub if it serves as a major transit node and otherwise it will be denoted an out-

port or simply a port. There may be multiple hubs in the service network. We shall generally assume that calls to multiple terminals within a single physical port can be accommodated by adding sufficient time for the port stay and thus limit the scope to considering only physical ports. This assumption will not restrict the generality of the model but serve to limit the network size.

In maritime liner terminology, routes are referred to as *rotations* where a rotation is usually defined as a sequence of port calls that form a closed cycle. Typically, a rotation will contain at least one port classified as a hub and by natural convention this hub will serve as the definition of the start and end point of the rotation. One completion of a rotation will be referred to as a voyage. The set of rotations operated form the backbone of the service network.

Products represent the interface toward the customers (shippers) and are defined as origin-destination freight services based on the set of rotations. Products may span multiple rotations thus requiring the transshipment of containers between different vessels. Although the concepts of consolidation and transshipment of containerized freight is central to many container liner operators we shall assume that transshipments are not allowed in the design of the master schedule. This assumption greatly simplifies the modeling and is justified by the fact that only around 2% of the total demand volume in the case considered in section 5.5 is transshipped and that transshipments are very costly and time consuming operations and thus undesirable. Additionally, since we are designing a master schedule, i.e., performing strategic planning, we are essentially determining the primary products that we want to offer and these generally do not include costly transshipments. Secondary products spanning two or more rotations can later be derived from the master schedule.

A fleet of *vessels* consisting of container ships is deployed to execute the planned rotations and although new vessels can be leased on a short term basis, a definition of the available fleet will influence the construction of rotations. Vessels will be divided into classes based on a certain set of characteristics including nominal capacity, ice-class and other physical properties such as draft or length that may restrict compatibility with certain ports. Vessel capacity is assumed expressed as total slot capacity measured in twenty-foot equivalent units (TEU) and corresponds to one standard container with a length of 20 feet. Associated with each vessel class is a cost determined by the time charter contract. This cost can be specified as a daily cost and together with the fuel cost this defines the cost of operating a vessel from a particular class. Individual vessels are deployed on a set of rotations such that the total fleet balances over a

period of time corresponding to the planning horizon.

The final component that forms the basis for the master schedule problem is a *demand* forecast detailed as a series of freight volumes specified on origin-destination pairs and time of availability. We assume demand to be independent of the service offered, known, and static within the planning horizon. However, the proposed modeling approach is capable of handling service level dependent demand. In the context of the specific case considered in this paper, the demand pattern is assumed to be recurring with a frequency of between one and three weeks and this will also be the basis for the length of the planning horizon. This assumption is not restrictive with respect to a typical planning process but may result in the failure to capture long term demand dynamics. We shall later discuss means to overcome this limitation.

Additional attributes such as the type and weight may also be associated with the individual demands. Forecast freight volumes will be required to be serviced in full in a valid master schedule and will drive the selection of the rotation set. This requirement combined with restrictions on the permitted delay of freight pickup allows for the implicit specification of service level requirements in terms of minimum visit frequency at individual ports during a predefined period of time. A further consequence of the “must carry all demand” requirement is that no considerations are given to revenue.

Based on the above discussion, the container feeder master schedule problem becomes that of selecting the minimum cost set of rotations that cover all forecast demand while satisfying topological constraints imposed on the individual rotations and the network as a whole.

5.3 Mathematical Model

The main idea behind the modeling approach adopted in this paper is to exploit the separability of the feeder service network design problem to decompose the problem into a series of smaller and independent problems. This leads to a structure where a master problem determines the routes that should be operated and how freight should be allocated to these routes. Two sub-problems dynamically update and augment the set of routes considered in the master problem and feasible freight allocations for this set of active routes respectively. Decomposing the problem in this fashion allows for a more even division of the overall problem complexity but also presents a few challenges that will be discussed in the following sections.

Before we present the master problem in section 5.3.2 the following section will introduce the network representation that provides the basis for the later discussions.

5.3.1 Service Network Representation

The foundation for the service network is a set of ports \mathcal{P} at which freight forwarding service must be offered. In this work, ports correspond to geographically distinct locations where vessels can load and discharge containers although this is not a limitation of the model. Ports can be divided into hub and non-hub ports. Hub ports $\mathcal{P}_H \subseteq \mathcal{P}$ define the interface toward other (deep-sea) carriers networks and also play a central role in the construction of services.

To capture the dynamics of the demand for freight forwarding service as well as the scheduling of vessels it is necessary to incorporate a temporal dimension into the flow of assets over the physical network. One approach to handling scheduling aspects in service network design problems is to expand the physical network with a time dimension leading to what is often referred to as a *space-time* network. This approach has its merits in problems where a discretization of the planning horizon is feasible and a meaningful choice of interval length exists.

Expanding the physical network to a planning horizon with $T + 1$ time periods we get a graph $G = (\mathcal{N}_v, \mathcal{A}_v)$ for each vessel class $v \in \mathcal{V}$ where each node $i \in \mathcal{N}_v$ is a tuple (p, t) representing a port $p \in \mathcal{P}$ at a period $t \in [0, 1, \dots, T]$. Note that nodes are only created at periods where a given port is open for service. Arcs are introduced between two nodes $i = (p_i, t_i)$ and $j = (p_j, t_j)$ if a vessel of the corresponding class can feasibly depart from p_i at time t_i and reach (including port service time at p_j) p_j at time t_j with p_j being open for service at t_j . Arcs that cross the end of the planning horizon, i.e., arcs (i, j) where $t_i + t_{ij} > T$ with t_{ij} being the transport time from i to j , will wrap around to the beginning of the planning horizon ($t_j = (t_i + t_{ij}) \bmod T$). In the real-world case considered in this work the planning horizon is of a length of 7 or 14 days and will be discretized into time intervals corresponding to the length of one day. Section 5.4.1 will discuss the role of vessel graphs in connection with rotation topology restrictions and rotation construction. Figure 5.1 illustrates the principles of the vessel graph on a network with seven time periods and four ports, one being a hub (H0). Each line corresponds to a feasible leg and the figure also illustrates the principle of a closed port (P0 at time 5) which is bypassed resulting in a later arrival at that port. To aid readability, only a subset of the feasible arcs have been included.

depends on the physical port-time instance $i = (p, t)$ and the vessel class.

With each vessel class v we associate a set of feasible routes \mathcal{R}_v , where the term route is used to generically represent a timed sequence of port calls that begin and end at a hub port. Routes may correspond to complete (maintain node balance across the planning horizon) or partial rotations. Each route $r \in \mathcal{R}_v$ consists of a set of legs $\mathcal{L}_r \subseteq \mathcal{A}_v$ connecting the ports serviced on that particular route, but at the master problem level, only the end points of the routes are considered with the purpose of maintaining vessel balance. Special restrictions on the vessel capacity u_{vij} for individual legs $(i, j) \in \mathcal{A}_r$ may exist.

An example of a network with six routes (solid lines) operating between four hubs over seven time periods is illustrated in Figure 5.2. Four routes balance to form a rotation with a period of twice the length of the planning horizon thus requiring two vessels to maintain a weekly frequency. In the context of the master problem a route need not be limited to a duration corresponding to the length of the planning horizon as vessel consumption is handled through the concept of a count line discussed later in this section.

The complete set of feasible routes for all vessel classes is denoted $\mathcal{R} = \bigcup_{v \in \mathcal{V}} \mathcal{R}_v$ with $\mathcal{R}_v(i)^+$ and $\mathcal{R}_v(i)^-$ denoting the set of routes from class v starting and ending in a node $i \in \mathcal{N}_H$ respectively. We introduce the decision variable $y_r \in \{0, 1\}$ to denote the decision to operate route r such that $y_r = 1$ if route r is operated and $y_r = 0$ otherwise. Associated with each route $r \in \mathcal{R}$ is a cost c_r which includes fixed time charter costs as well as variable operating costs.

Let \mathcal{K} define the set of commodities that must be serviced. Each individual commodity $k \in \mathcal{K}$ is defined by an origin $O(k) \in \mathcal{P}$ where the commodity becomes available, a destination $D(k) \in \mathcal{P}$ where the commodity must be delivered, and a volume $d^k > 0$ measured in TEU. Additionally, we associate a time of availability t_a^k and latest time of pickup t_l^k with each commodity $k \in \mathcal{K}$. Commodities thus correspond to demand forecasts aggregated according to the time discretization and the above dimensions.

To avoid handling capacity constraints in the master problem we introduce the concept of a commodity group $g \subseteq \mathcal{K}$ which corresponds to a set of commodities that can feasibly be serviced on a given route r . The set of all feasible commodity groups for a route r is denoted \mathcal{G}_r . Let $x_{gr} \in [0, 1]$ denote the fraction of commodity group g that is allocated to route $r \in \mathcal{R}$. Finally, let \mathcal{G}_r^k denote the set of commodity groups for route r that cover commodity $k \in \mathcal{K}$. Section 5.3.3 will present the concept of

commodity groups in further detail.

The feeder liner service network design (LSND) model is defined as follows

$$(\text{LSND}) \quad \min \sum_{r \in \mathcal{R}} c_r y_r + \sum_{v \in \mathcal{V}} \sum_{i \in \mathcal{N}_H} c_{vi} w_{vi+} \quad (5.1)$$

s.t.

$$(\pi^k) \quad \sum_{r \in \mathcal{R}} \sum_{g \in \mathcal{G}_r^k} x_{gr} = 1, \quad \forall k \in \mathcal{K} \quad (5.2)$$

$$(\alpha_r) \quad \sum_{g \in \mathcal{G}_r} x_{gr} \leq y_r, \quad \forall r \in \mathcal{R} \quad (5.3)$$

$$(\beta_{vi}) \quad \sum_{r \in \mathcal{R}_v(i)^+} y_r + w_{vi+} - \sum_{r \in \mathcal{R}_v(i)^-} y_r - w_{vi-} = 0 \quad \forall v \in \mathcal{V}, i \in \mathcal{N}_H \quad (5.4)$$

$$(\gamma_v) \quad \sum_{r \in \mathcal{R}_v} \delta_r y_r + \sum_{i: w_{vi+} \in \delta(\mathcal{W}_v)} w_{vi+} \leq n_v \quad \forall v \in \mathcal{V} \quad (5.5)$$

$$x_{gr} \in [0, 1], \quad \forall r \in \mathcal{R}, g \in \mathcal{G}_r \quad (5.6)$$

$$y_r \in \{0, 1\}, \quad \forall r \in \mathcal{R} \quad (5.7)$$

$$w_{vi+} \in \mathbb{Z}_+, \quad v \in \mathcal{V}, \forall i \in \mathcal{N}_H. \quad (5.8)$$

The objective function (5.1) minimizes the sum of costs associated with operating selected routes and the cost of having vessels stay idle in ports between routes. Although vessels would generally not lay idle in ports but rather anchor outside the ports, idling is usually undesirable and may be penalized using the waiting variables. Furthermore, if deployed, a vessel incurs charter costs for the whole planning period, even if it is idle for a part of the period.

Commodity constraints (5.2) enforce that all commodities must be serviced in full while capacity constraints (5.3) ensure that sufficient routes are operated to satisfy forecast demand enabled by the selection of commodity groups.

Constraints (5.4) are asset flow balance constraints enforcing that the vessels from each fleet class balance across the planning horizon and the

individual ports. Note that congestion issues are not considered in this model but can be incorporated on hub level by introducing appropriate constraints. Finally, constraints (5.5) are count constraints that enforce that the number of vessels deployed from a particular fleet v does not exceed the fleet size. This is achieved through the principle of a count line which is simply a point between two time periods at which the number of route crossings δ_r , $r \in \mathcal{R}$ are counted. $\delta(\mathcal{W}_v)$ defines the set of waiting variables of a particular vessel class $v \in \mathcal{V}$ crossing the count line.

Two of the main challenges associated with LSND are the construction of the sets \mathcal{R} and \mathcal{G}_r as even moderately sized problems will potentially yield a very large set of feasible routes and corresponding large sets of commodity groups for these routes. However, if we can intelligently select only a small subset of promising routes $\tilde{\mathcal{R}} \subset \mathcal{R}$, LSND will present a number of benefits including capturing complex route and commodity allocation rules in sub-problems and yielding an integer problem that is more easily solved than traditional compact network design formulations. A discussion about how to overcome the challenge associated with the set of routes through the dynamic update of the set of promising routes $\tilde{\mathcal{R}}$ is presented in section 5.4. The following section will discuss the augmentation of the sets \mathcal{G}_r , $r \in \tilde{\mathcal{R}}$ based on an assumption that the set $\tilde{\mathcal{R}}$ of promising routes is known.

5.3.3 Commodity Group Generation

The master problem (5.1)–(5.8) introduced the concept of a commodity group. Before we continue the discussion of the generation of commodity groups a presentation of some basic definitions is beneficial for the exposition.

We say that a commodity $k \in \mathcal{K}$ is feasible for a route $r \in \mathcal{R}_v$ from a vessel class $v \in \mathcal{V}$ only if the following conditions are satisfied:

- The route visits both $O(k)$ and $D(k)$ with the visit to $O(k)$ before $D(k)$.
- The route arrives at $O(k)$ a time $t_{rO(k)}$ such that $t_a^k \leq t_{O(k)} \leq t_l^k$.
- All legs $(i, j) \in \mathcal{A}_r$ of the route r between $O(k)$ and $D(k)$ satisfy $u_{vij} \geq d^k$.

Note that satisfying the above conditions does not guarantee that commodity k will be serviced by vessel route r in a particular solution since other commodities may be assigned to the same route thus consuming capacity. The set of feasible commodities for a particular route r will be denoted $\mathcal{K}_r \subseteq \mathcal{K}$. Partially implied in the last condition is the re-

quirement that a commodity, if assigned to a commodity group, must be assigned in full to that particular group.

A feasible commodity group $g \in \mathcal{G}_r$, $g \subseteq \mathcal{K}_r$ is a set of commodities that can be serviced on a given route $r \in \mathcal{R}_v$ while respecting the capacity constraint on each arc of the route. Depending on the number of commodities feasible for a particular route there may be a large number of feasible assignments of commodities to that route of which only a few may be attractive. It is undesirable and impractical to exhaustively enumerate all feasible commodity assignments. Instead, we define a commodity assignment problem which, based on dual information from the linear programming relaxation of LSND, allows us to dynamically generate attractive commodity assignments. In other words, we employ column generation to implicitly enumerate commodity assignments for the individual routes of LSND and replace \mathcal{G}_r with $\tilde{\mathcal{G}}_r \subseteq \mathcal{G}_r$ for each route.

Given a feasible commodity group $g \in \mathcal{G}_r$ the reduced cost \tilde{c}_g of g is defined as follows:

$$\tilde{c}_g = - \sum_{k \in g} \pi^k - \alpha_r, \quad (5.9)$$

where π^k are the duals corresponding to commodity cover constraints, i.e., constraint set (5.2), and α_r corresponds to the route capacity constraints (5.3). Let $x^k \in \{0, 1\}$ denote the decision to allocate a commodity to a vessel such that $x^k = 1$ if commodity $k \in \mathcal{K}_r$ is allocated to vessel class $v \in \mathcal{V}$ operating on route $r \in \mathcal{R}_v$ and $x^k = 0$ otherwise. Recall that each commodity has an associated origin and destination. Given a particular route r we define by $p_r(k)$ the path, i.e., the sub-sequence of legs, on which a commodity will occupy capacity if allocated.

With these definitions we are now able to formally state the commodity group generating (CGG) sub-problem associated with a particular route $r \in \mathcal{R}_v$, $v \in \mathcal{V}$.

$$(\text{CGG}_r) \quad \min - \sum_{k \in \mathcal{K}_r} \pi^k x^k \quad (5.10)$$

s.t.

$$\sum_{k \in \mathcal{K}_r: (i,j) \in p_r(k)} d^k x^k \leq u_{vij}, \quad \forall (i,j) \in \mathcal{A}_r \quad (5.11)$$

$$x^k \in \{0, 1\}, \quad \forall k \in \mathcal{K}_r \quad (5.12)$$

The objective (5.10) is to minimize the variable part of the reduced cost for a commodity allocation while respecting the capacity constraints on the individual legs of the route (5.11). We omit the dual contribution α_r from the cover constraint since this is a constant. However, in the case where $\pi^k \leq 0$ for all $k \in \mathcal{K}_r$, the resulting commodity group will cover no commodities but can still represent a negative reduced cost column if $\alpha_r > 0$. This can occur if it is more expensive to utilize port waiting arcs in the master problem than deploying a vessel on a route with no commodities allocated.

One important note is that the structure of the problem given in (5.10) – (5.12) depends on the route which means that we potentially have $|\mathcal{R}|$ sub-problems. This is generally not a desirable property but if we are able to limit the number of routes that are considered and dynamically create sub-problems as new routes are considered, the impact of having unique sub-problems for each route may be reduced. Furthermore, since the sub-problem is restricted to only consider commodities feasible for a single route, the sub-problem is relatively small and easily solved. An additional consequence of this construction is that an increase in the size of \mathcal{K} , i.e., the set of all commodities, only has a small impact on the performance of the commodity group generating sub-problems.

5.4 Solution Methodology

We pursue two different strategies for the solution of LSND. The first is based on enumeration of feasible route candidates which are added to a route pool and combined with a mechanism for selecting routes from this route pool. The second strategy iteratively augments a possibly empty route pool by building new candidate routes using dual information from linear programming relaxation of LSND (LP-LSND). Both strategies build on a principle of iteratively expanding the model both in terms of variables and constraints. Algorithmically, the two strategies share the same overall flow as outlined in Figure 5.3

The first step of the algorithm initializes the route pool with a set of route candidates. Next, we select a subset of routes from the route pool and add these to the LP-LSND. This resulting problem instance will be referred to as the restricted master problem r-LSND. The set of routes in the r-LSND is referred to as the active routes and denoted $\tilde{\mathcal{R}}$. Since the initial set of selected route candidates may not constitute a feasible solution to LP-r-LSND we introduce a feasibility column for each commodity $k \in \mathcal{K}$ with which we associate a high cost. Based on dual

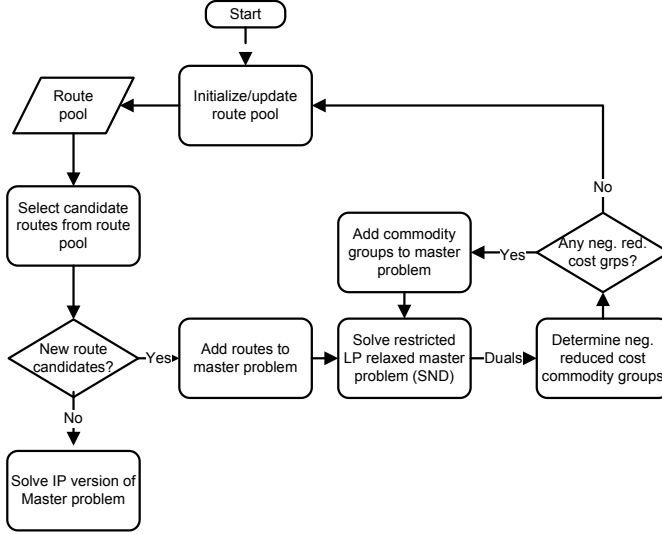


Figure 5.3: Algorithmic flow.

information from LP-r-LSND we then iteratively generate new commodity groups for all routes currently in the set of active routes $\tilde{\mathcal{R}}$, add these to LP-r-LSND, and re-solve LP-r-LSND. This process continues until no more negative reduced cost (objective improving) commodity groups can be found or the improvement in the objective value falls below a defined tolerance. The algorithm then returns to the main loop to update the route pool and subsequently select new route candidates to enter $\tilde{\mathcal{R}}$.

When no more route candidates to enter $\tilde{\mathcal{R}}$ are found or the improvement in the objective between two main loop iterations is below a threshold value ϵ_{obj} , r-LSND is solved using standard mixed integer program (MIP) solver software. Since we only add routes and commodity groups to the master problem in the root node of the branch and bound tree, we allow individual commodities to be covered by more than one route when solving the MIP, i.e., we change partitioning constraints (5.2) to cover constraints.

5.4.1 Route Topology and Enumeration

Business rules derived from the current operating environment as well as properties of acceptable routes serve to reduce the set of feasible routes. The following properties will be enforced when constructing route can-

didates.

- A route must begin and end at a hub port.
- A port is not called more than twice between visits to a hub. This generally translates to an outbound and an inbound call on any given route.
- Average export volumes combined with maximum vessel capacity and typical hub to out-port to hub demand pattern naturally limits the number of ports that are called on any route.
- Individual routes cannot be of a duration longer than the length of the planning horizon.

As mentioned in section 5.3.1, physical restrictions related to vessel-port compatibility are embedded in the individual vessel graphs. Furthermore, any operational restrictions imposed on service between certain port pairs are also captured in the vessel graph.

Route enumeration is performed using a label setting algorithm with an initial label for each node in \mathcal{N}_H and a route is considered complete and added to the route pool as soon as it reaches a second hub port. This means that a route only contains two hub ports; the first and last port in the sequence. Maintaining these minimal routes and balancing them in the master problem has the advantage of reducing the total number of feasible routes significantly.

Since the properties imposed on individual routes are only moderately restricting, we will generally expect to have many routes with very similar properties. This is undesirable from an algorithmic point of view and the many routes may not be equally attractive from a business perspective. To some extent, this issue can be overcome by introducing an approximate dominance criterion which will be discussed in section 5.4.3.1.

5.4.2 Route Selection and Elimination

Unfortunately, even with the above described topological constraints imposed on the set of feasible routes, the total number of routes is usually still too large to handle explicitly in the master problem. Thus, we need a mechanism for selecting new routes from the route pool to be admitted into the master problem as well as a mechanism for eliminating routes that no longer appear promising.

For the selection of entering routes we introduce a measure of the reduced cost estimate of a particular route beginning at hub node s and ending

at hub node t :

$$\bar{c}_r = c_r - \beta_{sv} + \beta_{tv} - \delta_r \gamma_v - \pi_g \quad (5.13)$$

where $\pi_g = \sum_{k \in g} \pi_k$ is an estimate for the reduced cost incurred by a feasible commodity group and serves as a substitute for the route capacity dual α_r which is only defined for routes in LP-r-LSND. The commodity group g is obtained by greedily adding commodities in order of decreasing dual value while respecting capacity constraints on each leg of the route. We note that π_g is a lower bound on the reduced cost contribution from the commodity cover constraints as the estimate is based on a greedy heuristic and a allocation may exist which results in a lower value of π_g . β_{sv} and β_{tv} are the dual contributions of the balancing constraints and γ_v is the dual corresponding to the fleet size constraint.

Obviously, the reduced cost estimate (5.13) does not capture the interaction between the multiple routes of the master problem and as a result it can be overly optimistic with respect to the attractiveness of a certain route when the duals of the LP-r-LSND are far from optimal. To avoid having the LP-r-LSND grow too fast we only admit the n_r routes with the most negative reduced cost estimate \bar{c}_r in each main loop iteration.

Elimination of routes is based on the reduced cost of routes currently in the master problem such that routes with a reduced cost \tilde{c}_r greater than some positive cut-off value are eliminated from the master problem. In practice, routes are only deactivated and can easily be re-activated if they become attractive at a later iteration of the algorithm. Active elimination of unpromising routes reduces the time that is spent in each main loop iteration of the algorithm but may lead to an increase in the number of iterations required to reach the termination criteria.

To account for the fact that the route selection criterion may be optimistic in the classification of route attractiveness and to reduce cycling we introduce a tabu list into which eliminated routes are inserted. Routes will remain in the tabu list for a predetermined number of iterations after which point they may again be admitted into the master problem. Computational experiments suggest that two main loop iterations is a reasonable choice for the length of the tabu period.

5.4.3 Greedy Heuristic

For larger real-world problems, exhaustively enumerating and maintaining a pool of all feasible routes becomes intractable and an alternative strategy for managing the route pool is needed. One such alternative

strategy is to initialize the pool with a small set of promising routes and then dynamically augment the route pool with new route candidates based on dual information from an LP optimal solution to the r-LSND.

To augment the route pool we employ a basic greedy sequential insertion heuristic. The heuristic begins with a possibly empty route and iteratively inserts commodities into the route using dual information in the evaluation of the cost of the insertion. This approach is similar to the greedy construction algorithm employed by Sol and Savelsbergh (1994) Sol and Savelsbergh (1994).

Analogous to the route selection criteria in section 5.4.2, the attractiveness of a new route is measured by the estimated reduced cost (5.13) with $c_r = \sum_{(i,j) \in \mathcal{A}_r} c_{ij}$ where c_{ij} is the cost of servicing leg (i, j) including port costs at i . However, $\pi_g = \sum_{k \in g} \pi^k$ now corresponds to the true dual contribution of the commodities serviced on the particular route using the corresponding commodity group g .

Let $\Delta \bar{c}_r^k = \Delta c_r^k - \pi^k$ denote the increase in the estimated reduced cost of route r by inserting commodity k at the position that results in the lowest increase in the reduced cost c_r of route $r \in \mathcal{R}$. Commodities that cannot feasibly be inserted into a route are assigned $\Delta \bar{c}_r^k = \infty$. At each iteration of the dual based greedy insertion heuristic, a commodity k where $k = \operatorname{argmin}_{k \in \mathcal{K}} \{\Delta \bar{c}_r^k\}$ is inserted at the corresponding lowest cost position if $\Delta \bar{c}_r^k < 0$. To account for the fact that a commodity may not always by itself result in decrease in the reduced cost, we accept a commodity k with $\Delta \bar{c}_r^k > 0$ into a route with a probability $p = \bar{p} e^{-\Delta \bar{c}_r^k / c_r}$. The heuristic terminates when no new commodities can be inserted into the route without increasing the estimated reduced cost or violating route constraints.

One of the benefits of the greedy insertion heuristic is that if a feasible route with a negative reduced cost estimate is found, a feasible commodity group is also found. Both will be added to LP-r-LSND and the route will also be added to the route pool. Furthermore, since intermediate routes are required to be feasible, any route with negative reduced cost estimate found during the heuristic search can be added to the route pool. Finally, the greedy heuristic can be run for a series of partial routes at each iteration allowing for the addition of multiple routes to the route pool.

5.4.3.1 Initializing the Route Pool

Several options for the initialization of the route pool exist. The first option is to base the initial routes on those currently operated by the logistics service provider. This option is obviously only available when the goal is to optimize an existing network. The second and most straightforward option is to fix appropriate initial values for the commodity duals in LP-r-LSND and use the greedy insertion heuristic to obtain an initial set of routes. Finally, we may seek to select a promising subset from the set of all feasible routes with which the route pool is initialized.

We formalize an intuitive understanding of comparative route attractiveness with the introduction of an approximate dominance criterion.

Definition 5.1 (Approximate Dominance) Given two routes r_1 and r_2 for vessel class v we say that r_1 approximately dominates r_2 if:

- r_1 and r_2 start and end at the same vertices of the vessel graph.
- The duration of r_1 is no longer than that of r_2 , i.e., $l(r_1) \leq l(r_2)$.
- r_1 covers a superset of the commodities covered by r_2 , i.e., $\mathcal{K}_{r_2} \subseteq \mathcal{K}_{r_1}$.
- $c_{r_1} < c_{r_2}$. □

Under certain conditions, Definition 5.1 may result in a route r_1 dominating r_2 even though r_2 is more attractive. This happens when the port call sequence combined with capacity constraints and the set of feasible commodities means that r_2 is able to service more commodities than r_1 . Despite this fact, computational experiments suggest that solutions based on a route pool only containing routes considered non-dominated according to Definition 5.1 are not far from solutions based on larger route pools.

5.5 Computational Experiments

In order to evaluate the feasibility of LSND in the context of a practical application, a series of planning scenarios are constructed. Each scenario is based on a real-world case with data from a container feeder operator servicing more than 25 ports in 12 different countries and moving more than one million TEUs annually with a fleet of around 30 vessels ranging in capacity from 500 to 1400 TEU. Individual scenarios correspond to typical planning scenarios investigated by the feeder operator on a regular basis. Characteristics of the individual scenarios have been summarized

in Table 5.1 together with the objective value of the best known MIP solution (bMIP). The best known solutions are obtained by running the different algorithms with the best parameter settings found during the experiments and a time limit of 12 hours. Finally, Table 5.1 also contains three smaller scenarios that will be used to compare and evaluate the performance of the two different solution strategies presented in section 5.4.

Instance	Vsl Class	Ports	Comm	Periods	bMIP
v3p5c42t7	3	5	42	7	1049.47
v3p5c63t7	3	5	63	7	1296.95
v1p6c90t7	1	6	90	7	2038.79
v2p6c90t7	2	5	90	7	1973.96
v3p6c90t7	3	5	90	7	1939.33
v3p7c112t7	3	7	112	7	2226.43
v4p9c96t7	4	9	96	7	1036.31
v5p12c271t7	5	12	271	7	4410.80
v6p14c307t7	6	14	307	7	4871.86
v6p16c325t7	6	16	325	7	5700.40
v6p10c355t14	6	10	355	14	24185.56

Table 5.1: Characteristics of the scenarios specified by number of vessel classes, physical ports, total number of O-D commodities (Comm), and time periods (Periods). Top three instances are synthetic comparison scenarios.

All calculations are performed on a computer with two Intel Xeon quad core processors running at 2.66 GHz with 16 GB of system memory and CPLEX 11.1 is employed as the LP and MIP solver. The software is not parallelized and thus only takes advantage of a single core. Reported running times are all wall clock time and deviations of 5-10% between runs of the same instance may be observed depending on the workload on the machine.

Algorithms all follow the overall flow outlined in Figure 5.3 and differ only by the method used to initialize and update the route pool. We distinguish between complete a priori enumeration of all feasible routes (CE), complete a priori enumeration with only non-dominated routes inserted into the route pool (CED), empty initial route pool with heuristic route pool augmentation (H), and finally partial a priori enumeration with heuristic augmentation (PEH). With this notation, the route pool e.g. initialized by complete a priori enumeration will be denoted \mathcal{R}_P^{CE} using subscript P to denote pool and distinguish from the sets of feasible and active routes. In all experiments, commodity groups are added

by means of delayed column generation and only in the root node of the branch-and-bound tree.

5.5.1 Baseline

To set a frame of reference for the later computational experiments on the above outlined scenarios, we first try to establish the strength of the formulation LSND measured in terms of the relative gap between the root node bound LP-LSND and the best integer solution. To achieve this, we generate a set of small baseline scenarios for which we can exhaustively enumerate all feasible routes that are all initialized as active, i.e., $\tilde{\mathcal{R}} = \mathcal{R} = \mathcal{R}_P^{CE}$. We then continue to generate commodity groups and add these to the master problem until no new negative reduced cost commodity groups can be found. The objective value of the LP-LSND is then a true lower bound on the optimal solution to each baseline scenario.

Instance	N rts	Complete enum				Dominance removal			
		Root	bMIP	Gap		N rts	Root	bMIP	Gap
v3p5c42t7	4176	1005.21	1049.47	4.22%		3556	1005.21	1062.61	5.40%
v3p5c63t7	3738	1235.20	1297.31	4.79%		3250	1235.45	1297.31	4.77%
v1p6c90t7	5586	1896.50	2042.52	7.15%		4830	1896.59	2046.66	7.23%

Table 5.2: Root node lower bounds (Root), best MIP bound and relative gap (Gap) for the baseline scenarios with and without dominated routes.

Table 5.2 summarizes the lower bounds (Root) obtained by enumeration of all routes as well as the relative gap between the lower bound and the objective value of the best MIP solution (bMIP). The table also shows the number of routes in $\tilde{\mathcal{R}}$ (N rts) where $\tilde{\mathcal{R}} = \mathcal{R}_P^{CE}$ and $\tilde{\mathcal{R}} = \mathcal{R}_P^{CED}$ for the first and second column groups respectively. Bold indicates best known solution.

As we only generate commodity groups in the root node of the branch-and-bound tree, the best MIP bounds indicated in Table 5.2 are not guaranteed to be optimal. However, based on the lower bound we can observe that MIP bounds are within 7.2% of the optimal value.

Removing routes dominated according to Definition 5.1 from the route pool reduces the number of routes by between 13% and 15% and reduces execution time by between 19% and 28% while only marginally impacting the lower bounds and resulting in increases in best MIP objectives of up to 1.3%. These results suggest that the proposed approximate dominance criterion is indeed successful at eliminating unpromising routes while

maintaining diversity among routes in the reduced route set. The next section will expand on these observations with a discussion about the impact on the resulting MIP problems by using dominance.

5.5.2 A Priori Route Pool Initialization

The main idea behind the algorithm based on a priori route pool initialization with dual based route selection is to provide a large and possible complete set of feasible routes from which promising candidates can be selected. Discrimination and selection of route candidates to enter $\tilde{\mathcal{R}}$ is performed using the reduced cost estimate defined in (5.13).

For the scenarios investigated in this section, only the topological constraints described in section 5.4.1 are imposed on the routes generated and populated into the route pool. This also means that for the larger scenarios, complete enumeration of the feasible routes was not tractable within the constraints of the system memory available to the software. Thus, experiments are limited to the subset of scenarios presented in Table 5.3.

Prior to conducting experiments that form the basis for the comparison of the different methods for route pool initialization, we need to determine an appropriate value of n_r , i.e., the maximum number of routes selected to enter $\tilde{\mathcal{R}}$ at each iteration of the main loop. This parameter represents the compromise between limiting the model size and ensuring that a sufficient set of routes is available when solving the final r-LSND. Small values of n_r means fewer commodity groups will be generated at each iteration and that the dual information used to select routes from the pool is updated more frequently. However, since the dual based route selection criterion tends to favor routes with very similar properties, small values of n_r tends to result in a lower degree of variation in characteristics of the active routes. In contrast, large values of n_r tends to result in more routes being activated and more commodity groups being generated overall but with less groups per route resulting in larger gaps between the root bound and the best MIP objective.

Experiments have been conducted with $n_r \in \{10, 30, 50\}$. The algorithm performed consistently better for $n_r = 30$ in terms of MIP solution quality despite being terminated by the time limit on the MIP for two more scenarios than in the case with $n_r = 10$. Thus, for the remaining runs of the dual based route selection algorithm, the route increment has been fixed to $n_r = 30$.

We have fixed a time limit (T_{MIP}) of 900 seconds to solve the MIP

(r-LSND). In many of the scenarios, a good integer feasible solution to r-LSND is found relatively early in the branch-and-bound tree search justifying limiting the solution time. For all scenarios in this section, the root node was solved in less than 250 seconds.

Instance	Rt pool	Active	Root	bMIP	Qual
v4p9c96t7	700472	1095	927.26	1045.75 ¹	1.01
v3p7c112t7	81934	691	2004.80	2318.00 ¹	1.04
v3p6c90t7	22422	638	1826.04	1980.90 ¹	1.02
v2p6c90t7	14004	443	1851.80	1988.28	1.01
v3p5c42t7	4176	338	1014.03	1049.47	1.00
v3p5c63t7	3738	393	1238.58	1296.95	1.00
v1p6c90t7	5586	277	1899.99	2038.79	1.00

Table 5.3: Details of the results obtained from running the dual based route selection algorithm with route enumeration. Quality (Qual) is relative to the best known objective value. $n_r = 30$, $T_{MIP} = 900$.
¹Terminated due to time limit.

Table 5.3 summarizes the results obtained from running the dual based route selection algorithm with \mathcal{R}_P^{CE} on seven different scenarios. For each scenario, we indicate the total number of route candidates populated into the route pool (Rt pool), the set of routes in the final r-LSND (Active) as well as the root bound (Root), and the objective value of the best integer feasible solution found (bMIP). Finally, we indicate the quality (Qual) of the solutions relative to the best known solution.

Overall, the dual based route selection algorithm performs very well and generally produces solutions of good quality within the fixed time limit. Average quality is 1.01 and the algorithm contributes with two best known solutions. The results suggest that the reduced cost estimate (5.13) is in fact quite successful at identifying objective improving routes and that these routes also prove attractive in the integer feasible solutions.

For the larger scenarios we observe slightly higher root–MIP gaps than those seen in the baseline experiments. This is particularly evident for the scenarios where the MIP search was terminated due to the time limit. Allowing the MIP solver to run for a longer period of time closes the gaps to approximately 10%.

For the three baseline scenarios, solutions were equal to or better than those found using the complete route enumeration. This is attributed in part to the fact that one baseline problem was terminated due to time limit and in part to the fact that more commodity groups and thus

more feasible route packings are generated for each active route in the dual based route selection algorithm than with all feasible routes active from the start. An interpretation of the results in terms of root bound strength, indicates that the dynamically constructed r-LSND is better described by its corresponding LP relaxation (LP-r-LSND) than is the case for the r-LSND with all routes active ($\tilde{\mathcal{R}} = \mathcal{R}_P^{CE}$). This is reflected in the root–MIP gaps which are smaller for the dynamically constructed r-LSND than for the r-LSND with all routes active.

5.5.2.1 Removing Dominated Routes

To further analyze the effect of removing dominated routes, we again solve the seven scenarios discussed above but remove the routes dominated according to Definition 5.1 from the route pool.

Instance	Rt pool	Active	Root	bMIP	Qual
v4p9c96t7	102508	914	931.56	1086.10 ¹	1.05
v3p7c112t7	63417	589	2010.83	2277.76 ¹	1.02
v3p6c90t7	19097	659	1827.21	1939.33	1.00
v2p6c90t7	11958	445	1850.69	1984.24	1.01
v3p5c42t7	3556	311	1014.03	1087.64	1.04
v3p5c63t7	3250	474	1237.28	1304.93	1.01
v1p6c90t7	4830	289	1904.99	2065.52 ¹	1.01

Table 5.4: Results obtained by removing dominated routes from the route pools utilized to obtain the results in Table 5.3. ¹Terminated due to time limit.

The results in Table 5.4 exhibit behavior which is similar to that observed when employing dominance in the baseline scenarios. We see reductions in the total running time between 7% and 69% although for two cases the time increased by 29% and 106% respectively. In terms of solution quality however, the results are less consistent. In three of the scenarios, solutions found are better than those obtained using the full route pool (Table 5.3) but in the remaining cases the best MIP solution increases by up to 3.6%. Overall, however, the removal of dominated routes from the route pool can prove useful to reduce overall problem size when using enumeration to populate the route pool.

5.5.3 Heuristic Route Generation

Continuing the series of experiments, we now look at the two versions of the algorithm where the route pool is dynamically augmented with routes during the solution of the root node, i.e., the cases with route pools \mathcal{R}_P^H and \mathcal{R}_P^{PEH}

One additional parameter is introduced by the heuristic route generator; the maximum probability \bar{p} for accepting a commodity with $\Delta \bar{c}_r^k \geq 0$. For the results presented in this section we set $\bar{p} = 0.4$ which appears to provide a good compromise between constructing routes with $\bar{c}_r > 0$ and rejecting potentially promising commodity inserts.

The algorithm is initialized with an empty route pool and experiments are conducted for $n_r \in \{10, 30, 50, 70\}$. The overall best results were obtained using $n_r = 50$ and the results using this setting are presented in Table 5.5. As before, we fix a time limit on the MIP problem of $T_{MIP} = 900$.

Instance	Rt pool	Active	Root	bMIP	Qual
v4p9c96t7	490	466	962.98	1200.62	1.16
v6p10c355t14	1099	1096	21872.83	25248.66 ¹	1.09
v6p14c307t7	945	935	4085.24	5796.82 ¹	1.19
v5p12c271t7	634	625	3877.84	5044.14 ¹	1.14
v6p16c325t7	1161	1135	4557.34	6612.85 ¹	1.17
v3p7c112t7	336	330	2161.68	2727.42 ¹	1.23
v3p6c90t7	252	252	1971.92	2136.61	1.10
v2p6c90t7	222	222	2053.43	2424.17	1.23
v3p5c42t7	124	124	1070.12	1171.02	1.12
v3p5c63t7	245	245	1320.44	1462.71	1.13
v1p6c90t7	154	154	1970.14	2051.08	1.01

Table 5.5: Details of the results obtained from running the algorithm with heuristic route pool augmentation. $n_r = 50$, $T_{MIP} = 900$.

¹Terminated due to time limit.

When comparing to the results obtained by using a priori route pool initialization with dual route selection, it is evident that the heuristic does not achieve similar good solutions. The average quality (Qual) for the same seven scenarios as those considered in Table 5.3 is 1.14 compared to 1.01 for the dual based route selection with a priori route pool initialization (\mathcal{R}_P^{CE}).

Although results for the heuristic route pool augmentation are not on par with the results obtained using a more extensive a priori route pool

initialization they can still prove valuable to the problem owner. This is evidenced by the consistently high capacity utilizations that are achieved by the heuristic based algorithm. Average capacity (TEU) utilization is between 71% and 90% with the median lying between 78% and 93%.

Adopting a hybrid approach where the route pool is initialized with a smaller set of routes (\mathcal{R}_P^{PEH}), we obtain an average solution quality of 1.03 for the 11 scenarios in Table 5.5 and 1.02 when excluding the baseline scenarios. The same two numbers when using \mathcal{R}_P^H are 1.14 and 1.16 with and without the baseline scenarios respectively. This clearly shows the benefit of initializing the route pool with a limited set of shorter routes and then employing the heuristic to augment the pool with longer routes. The cost of this approach is a larger set of active routes $\tilde{\mathcal{R}}$ and resulting longer running times.

Experiments with a variation of the algorithm have been performed in an attempt to reduce the amount of time spent solving LP-r-LSND in each main loop iteration. It turns out that the extra time spent solving the LP-r-LSND between each call to the greedy heuristic for each vessel class to obtain updated duals results in better solutions than when only solving LP-r-LSND once for each main loop iteration. This observation also applies more generally where frequent updates to the duals between model modifications (additions of routes and commodity groups) yields faster convergence in the solution of the root node. On the other hand, as we only generate commodity groups and routes in the root node of the branch-and-bound tree, the performance of the algorithm also depends to some extent on using “intermediate” duals to obtain a sufficiently large set of routes and commodity groups. This observation supports the previous discussion regarding the settings for the value of n_r which also affects the frequency at which duals are updated.

5.6 Concluding Remarks

Inspired by a real-world case, we have presented a service network design model and solution method that, based on characteristics of the problem, exploit separability and allows us to develop an algorithm that dynamically extends the problem using dual information in two classes of sub-problems. The proposed model implicitly captures service frequency requirements and the decomposition enables the inclusion of additional restrictions on demand loads as well as route topology.

Computational experiments showed that the presented algorithms were capable of solving real-world problems of realistic size within a time

frame reasonable to the application of tactical and strategic planning. Measured by business defined metrics such as capacity utilization, all algorithms produce good results and can provide planners with a tool for quick evaluation of predefined operating scenarios. The concept of a route pool allows planners to provide an initial set of route alternatives to the algorithms with the option of having this set of routes expanded by the algorithms.

Experiments also showed that although large route pools provided the best results, initializing the route pool with a smaller set of feasible routes and then augmenting this pool using a greedy heuristic provided good results and enabled the solution of larger scenarios. Furthermore, the reduced cost estimate proved successful at identifying good subsets of routes even from large route pools.

In terms of improving the performance of the heuristic route pool augmentation algorithm, it would be interesting to implement additional routines for the construction of new route candidates. One option is to combine the greedy heuristic with a local search mechanism to extend the set of feasible routes examined. Funke et al. Funker et al. (2005) provide an extensive review of the use of local search methods for vehicle routing and scheduling problems, many of which will also apply to our problem. Finally, in addition to heuristics, it may prove beneficial to implement a dynamic programming algorithm which would be employed when no more route candidates are identified by the heuristics.

Bibliography

- R. Agarwal and E. Özlem. Ship scheduling and network design for cargo routing in liner shipping. *Transportation Science*, 42(2):175–196, 2008.
- R. K. Ahuja, T. L. Magnanti, J. B. Orlin, and M. Reddy. Applications of network optimization. In M. O. Ball, T. L. Magnanti, C. L. Monma, and G. L. Nemhauser, editors, *Network Models*, volume 7 of *Handbook in Operations Research and Management Science*, chapter 1. Elsevier, 1995.
- F. Broeze. *The Globalisation of the Oceans: Containerisation from the 1950s to the Present*. Number 23 in Research in Maritime History. International Maritime Economic History Association, 2002.
- M. Christiansen, K. Fagerholt, and D. Ronen. Ship routing and scheduling: Status and perspectives. *Transportation Science*, 38(1):1–18, 2004. ISSN 00411655.

- B. Funker, T. Grünert, and S. Irnich. Local search for vehicle routing and scheduling problems: Review and conceptual integration. *Journal of Heuristics*, 11:267–306, 2005.
- N. Hingorani, D. Moore, and K. Tornqvist. Setting a new course in the container shipping industry. Technical report, IBM Business Consulting Services, IBM Global Services, Route 100, Somers, NY 10589, U.S.A., 2005.
- D. I. Jaramillo and A. N. Perakis. Fleet deployment optimization for liner shipping, part 1. *Maritime Policy and Management*, 18(4):235–262, 1991.
- M. Levinson. *The Box*. Princeton University Press, 1st paperback printing edition, 2006.
- A. N. Perakis and D. I. Jaramillo. Fleet deployment optimization for liner shipping, part 1. *Maritime Policy and Management*, 18(3):183–200, 1991.
- B. J. Powell and A. N. Perakis. Fleet deployment optimization for liner shipping: an integer programming model. *Maritime Policy and Management*, 24(2):pp. 183–192, 1997.
- D. Ronen. Cargo ships routing and scheduling: Survey of models and problems. *European Journal of Operational Research*, 12(2):pp. 119–126, 1983.
- D. Ronen. Ship scheduling: The last decade. *European Journal of Operational Research*, 71(3):pp. 325–333, 1993.
- M. M. Sigurd, N. L. Ulstein, B. Nygreen, and D. M. Ryan. Ship scheduling with recurring visits and visit separation requirements. In G. Desaulniers, J. Derosiers, and M. M. Solomon, editors, *Column Generation*, GERAD 25th anniversary series, chapter 8, pages 225–245. Springer Science, 2005.
- M. Sol and M. W. P. Savelsbergh. A branch-and-price algorithm for the pickup and delivery problem with time windows. Technical report, Dept. of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands, 1994.
- P. Toth and D. Vigo, editors. *The Vehicle Routing Problem*. Discrete Mathematics and Applications. SIAM, 2002.
- UNCTAD. Review of maritime transport 2008. Technical report, United Nations Conference on Trade and Development, Geneva, 2008.

- F. Weldon, editor. *Operational Simulation of a Freighter Fleet*, Research Techniques in Maritime Transportation, Washington, 1959. National Research Council publication 720.
- F. L. Weldon. Cargo containerization in the west coast-hawaiian trade. *Operations Research*, 6(5):649–670, 1958.

CHAPTER 6

Paper 2: Network Transition in a Liner Container Shipping Application

Martin W. Andersen,
Department of Transport,
Technical University of Denmark

Oli B.G. Madsen,
Department of Transport,
Technical University of Denmark

Allan Olsen,
Department of Transport,
Technical University of Denmark

Abstract

Service network design is a central planning process in many freight forwarding businesses and an important determinant for the success of a company. While much work has addressed the design of efficient service networks, only limited attention has been given to the actual implementation of a new service network. This paper introduces a new problem, the network transition problem, which addresses the process of moving assets from operating an existing service network to a new adjusted service network. The network transition problem is described in the context of liner container shipping but finds application in other types of service networks.

We develop a parallel cooperative adaptive large neighborhood search heuristic employing both well known and problem specific neighborhood operators. The heuristic is analyzed using a series of synthetic test instances of various sizes and with attributes similar to those of the real-world case problem that inspired this work. Computational experiments suggest that the heuristic is robust against different instance characteristics and capable of solving large problem instances (up to 400 request) within a time reasonable for the application. A concrete case from a feeder company is used to evaluate the applicability of the approach in a real-world scenario. Results show that the heuristic is capable of producing solutions that are competitive to those produced by the problem owner measured in terms of traditional performance indicators such as capacity utilization.

Keywords: liner shipping, service network design, recovery, large neighborhood search

6.1 Introduction

Designing an efficient and reliable service network that meets customer requirements for freight forwarding services at the lowest possible cost is a central planning problem in many freight forwarding businesses. The design of the service network is a major determinant for the success of a company (carrier) and its competitive position in the market for freight forwarding services. At the same time it represents one of the most complex planning problems faced by carriers as it integrates multidimensional decisions with network wide correlations. This also means that it is usually not desirable or efficient to perform many minor

redesigns of the service network as this can lead to side effects in the network that are difficult to evaluate and may disrupt operations. However, sometimes it is necessary to adopt a local or regional approach to modifying the service network due to the complexity of migrating from the existing and active service network (with an associated schedule) and to the new network design (and new schedule) proposed by either manual planning or integrated service network design (SND) models such as the one proposed by Andersen et al. (2009).

In certain service networks such as those providing service by means of trucks or aircraft, operation is typically halted during the night. This allows for the repositioning of assets (trucks or aircraft) during the non-operational hours, thus providing the flexibility to essentially effectuate deployment of a new service network design over night. Such repositioning flexibility translates into a larger degree of freedom in the frequency and scope of service network redesigns. In contrast to these two modes of transport, liner based service networks such as container shipping, the subject of this paper, are an example of a transport mode in which vessels operate around the clock thus allowing no time for repositioning. Add to this the fact that repositioning a vessel can take days or even weeks and that freight forwarding obligations must be met during the period of repositioning. Furthermore, due to the nature of liner container shipping, ships never really become empty at any point in the network which further complicates the process of repositioning.

This paper addresses the problem of transitioning a fleet of vessels from operating one liner based service network to being deployed on another service network while meeting pre-existing freight forwarding obligations during the period of the transition. We denote this problem the *network transition* problem.

The network transition problem can be considered an extension of the classical pickup and delivery problem with time windows to include multiple alternative delivery locations for each commodity, possibly different start and end depots for each vehicle and the possibility of transferring freight to external vessels. The relationship will become more evident in section 6.2.

6.1.1 Related Work

To the best knowledge of the authors, no work has previously been published on the network transition problem. However, as mentioned previously, the problem shares many features with the pickup and delivery problem and certain schedule recovery problems.

Ball et al. (2007) provides a recent survey and overview of schedule recovery in air traffic problems and presents an aircraft recovery model which, although quite general and with more complex side constraints than those of the network transition problem, captures many of the properties of the network transition problem. Specifically, the model includes end requirements ensuring that the positioning of aircraft at the end of the recovery period is compatible with the following schedule. An extension of this model which includes additional congestion constraints is presented by Rosenberger et al. (2003) who propose a heuristic to reduce the size of the mixed integer aircraft recovery model.

With respect to vehicle routing and scheduling, much work has been done, both in terms of practical applications and more theoretical works. Some of the classical problems such as the vehicle routing and scheduling problem (Toth and Vigo (2002)) and its variants such as the pickup and delivery problem (PDP) (see Parragh et al. (2008a,b) and Berbeglia et al. (2007) for recent surveys) have been the basis for a large variety of algorithmic ideas and applied works. These problem classes typically assume some notion of a depot at which assets are available or alternatively allows repositioning of the assets. In businesses like the maritime based liner container freight transportation, repositioning assets can take several days during which time contractual freight must still be carried. This means that it is generally not feasible to perform repositioning with empty vessels adding to the complexity of operation during periods of network transition. A few works (Mitrović-Minić and Laporte, 2006, Cortés et al., 2010) consider variants of the PDP in which transshipment of freight between vehicles is allowed at pre-determined transshipment points. Only results from very small instances (six requests) are reported in Cortés et al. (2010) who use a branch-and-cut approach, while Mitrović-Minić and Laporte (2006) report solving problems with up to 100 requests using a two-phased heuristic.

Additional background on maritime planning problems can be found in two recent surveys by Christiansen et al. (2004, 2007). Furthermore, Crainic (2000) provides a general introduction to service network design in freight transportation and Andersen et al. (2009) discuss a concrete service network design problem in a liner container feeder application. This problem is closely related to the network transition problem from a business perspective as it supports the process of designing a new service network which in turn justifies and provides input to the network transition problem.

6.1.2 Contributions and Outline

The contributions of this paper are two-fold. First, we introduce a new problem, the network transition problem, addressing one of the key challenges of deploying new service network designs in existing and active networks. Second, we develop a general cooperative adaptive large neighborhood search (ALNS) framework and present a series of neighborhood operators that are applicable for the network transition problem. We develop a new set of synthetic test instances with which the qualities and performance of the ALNS framework are assessed and show the tractability of the solution approach in connection with instances of realistic size. Finally, we evaluate the behaviour of the ALNS in a real-world scenario using data from a feeder shipping company.

The following section will present the network transition problem in further detail. This section will also briefly discuss the related problem of network recovery. Section 6.3 will provide details on the solution approach which is based on a parallel cooperative adaptive large neighborhood search heuristic embedded in a simulated annealing framework. Finally, results from experimental work on a series of synthetic test instances as well as a real-world case will be reported in section 6.4 before concluding remarks are given in section 6.5.

6.2 The Network Transition Problem

The network transition problem will be described in the context of a specific case for a liner container feeder company but will find general application in liner based networks.

Three to four times a year, the feeder company will revise its master schedule in response to changing market conditions primarily determined by demand and new product requirements. In-between these network revisions, minor adjustments to the schedule may be performed as a result of changes in the schedules of the interfacing deep-sea liner operators. These minor adjustments are important as the deep-sea operators provide a large part of the total freight volume.

We shall refer to the system consisting of a fleet of vessels \mathcal{V} operating a schedule servicing a set of ports (locations) \mathcal{L} as the service network. A network transition may involve part of or the whole fleet of vessels depending on the scope of the service network change. During the transition from one service network to another the liner operator must meet previously engaged freight forwarding obligations while seeking to mini-

mize the cost of this transition.

The duration of the transition period will be decided by a network planner and is generally determined by tactical considerations as well as possible contractual obligations, both in terms of freight forwarding and vessel chartering. We shall assume that start and end time of the transition period is given. However, it is possible to adapt the solution approach described in section 6.3 to allow dynamic expansion of the transition period.

We partition the service network into two parts based on the transition status of the vessels servicing each of these two parts. The partial network formed by the vessels that are not transitioning \mathcal{V}_F and thus continue to operate the planned and *fixed schedule* represent the first part. These vessels will perform port visits during the transition period which are not changed. The fixed part of the service network will serve as a capacity resource to which cargo can be transshipped. As such, these vessels will not be explicitly considered in the model but rather serve as input to determine feasible transshipment locations as described later in this section. Vessels in \mathcal{V}_F may either be vessels from the fleet of the liner operator or they may be vessels from another operator with whom the liner operator has a vessel sharing agreement.

The second part of the service network is formed by the *transitioning* vessels \mathcal{V}_T and consists of two components distinguished by the status of their respective schedules. The first component is comprised of the *currently scheduled* port visits of the transitioning vessels. Each vessel involved in the transition will continue to operate its original schedule until the final port call immediately after the start of the transition period. The second component is comprised of the *new schedule* for the transitioning vessels. The new schedule for each individual vessel $v \in \mathcal{V}_T$ is represented by a single port visit prior to the end of the transition period T_{end} and a series of port calls following T_{end} .

Figure 6.1 illustrates a possible transition schedule for a set of transitioning vessels $\{A, B, D\}$ as well as a non-transitioning vessel C . Solid lines indicate the planned and fixed schedule with port visits indicated by line-end markers. Port visits for transitioning vessels to the left of T_{start} belong to the old schedule while port visits to the right of T_{end} belong to the new schedule. Dashed lines represent a series of possible port visits as they could look for a transition schedule being operated between T_{start} and T_{end} . Ultimately, the goal of the network transition problem is to determine this transition schedule.

For each transitioning vessel $v \in \mathcal{V}_T$ we define a start location l_s^v and an

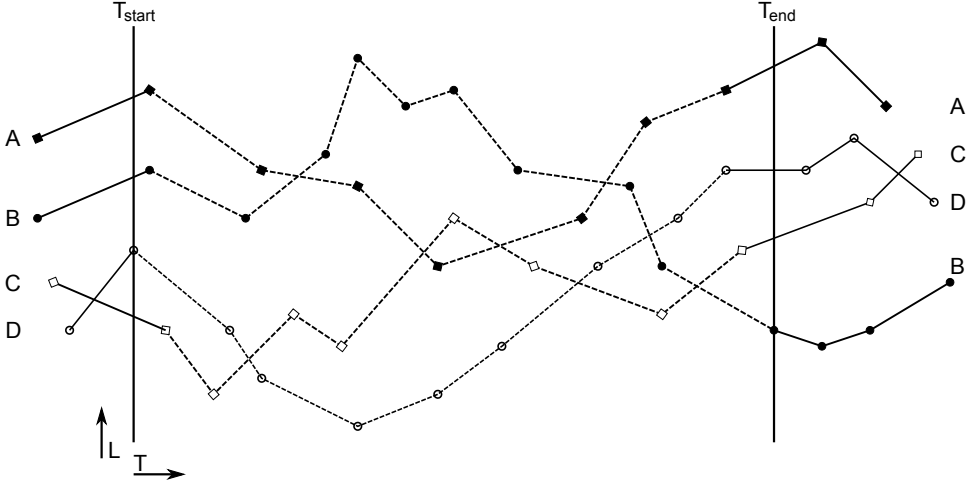


Figure 6.1: Example of a network transition with calls before T_{start} /after T_{end} belonging to the current and new schedule respectively. Solid lines indicate fixed schedules (port visits). A possible transition schedule is indicated by dashed lines.

end location l_e^v which may be geographically distinct. The start location is given by the last port visit from the old schedule immediately after T_{start} and similarly the end location is defined as the first port visit from the new schedule immediately prior to T_{end} . The start and end locations of a vessel have fixed time windows and constitute the fixed end-points of a vessel route. In terms of the real-world problem, fixed start and end locations translate into the requirement that each vessel $v \in \mathcal{V}_T$ must transition to a specific new schedule. Additionally, with each vessel $v \in \mathcal{V}_T$ we associate a capacity, a cruising speed, a cost per unit distance (nautical mile), and a canal fee payed on certain links. As there is a one-to-one correspondance between a route and a vessel, we shall also use \mathcal{V}_T to denote the set of individual vessel routes.

The demand for freight forwarding service that must be met during the transition period is represented by a set of commodities \mathcal{K} . We will not distinguish explicitly between contractual and spot cargo but rather assume that all demand must be met during the transition period. Non-serviced commodities will incur a penalty to allow infeasible solutions during the search procedures described in section 6.3. With each commodity $k \in \mathcal{K}$ we associate a freight amount d^k as well as a pickup location $l_P^k \in \mathcal{L}$ and a primary delivery location $l_{D_1}^k \in \mathcal{L}$. In addition to

the primary delivery location, we define a set of secondary delivery locations $\hat{\mathcal{L}}_D^k$ for each commodity $k \in \mathcal{K}$. These secondary delivery locations are obtained through a mapping based on the schedules of the vessels $v \in \mathcal{V}_F$ as well as the new schedules for vessels $v \in \mathcal{V}_T$.

In general, the feasible mappings of secondary delivery locations are determined by the schedules of the vessels involved in the transition problem, i.e., $v \in \mathcal{V} = \mathcal{V}_F \cup \mathcal{V}_T$, together with products that the company wishes to offer. Here, a product specifies how an origin-destination demand can be satisfied through the combination of two or more routes and a set of specified transshipment locations. When secondary delivery locations imply transshipment to a vessel $v \in \mathcal{V}_F$ we shall generally assume that sufficient capacity is available on this vessel to meet transshipping demand. Furthermore, we do not allow transshipment of freight between vessels $v \in \mathcal{V}_T$. Only secondary delivery locations for which final delivery to the primary destination can be achieved within the primary time window will be considered feasible.

With each commodity $k \in \mathcal{K}$, we associate a base profit p^k which will be earned when that commodity is delivered at its primary delivery location. Since the mapped secondary delivery locations may involve transshipments, the profit obtained from servicing a commodity will depend on the chosen location of delivery. The highest profit is achieved by delivering to the primary location while secondary delivery locations can incur transshipment costs that may depend on the location and thus result in a lower profit. Example: We may pickup a commodity k on vessel v at location l_P^k during the transition period for final delivery at location $l_{D_1}^k$. However, rather than having to deliver the commodity within the transition period, we may decide to leave it on the vessel v given that the vessel visits location $l_{D_1}^k$ in its new schedule (after the end of the transition period). This would effectively result in the mapping of a delivery location corresponding to the first port visit of vessel v in the new schedule. In this example, no transshipment cost is incurred as the commodity remains on vessel v . Secondary delivery locations that result in a negative profit will not be included in $\hat{\mathcal{L}}_D^k$.

For the remainder of this paper, the term request will be used to denote a commodity together with any mapped secondary delivery locations and again use the symbol \mathcal{K} to denote the set of all requests. All locations (pickup and delivery) associated with requests have associated time windows defining when a request can feasibly be handled (picked up or delivered). These time windows can be used to specify maximum lead time restrictions between pickup and delivery of a request thereby also implicitly specifying service level on specific origin-destination com-

binations.

With the above definitions, the objective of the network transition problem becomes that of determining the lowest cost sequence of visits that take the vessels in \mathcal{V}_T from their old schedule to their new schedule while servicing the available demand. The cost components of the objective consist of expenses to bunker (fuel), canal fees, and port fees. Additionally, request profits contribute with negative costs.

6.2.1 Network Recovery

It is relevant to note that although this work focuses on the network transition problem, the solution approach is equally capable of addressing the closely related problem of network recovery. This problem has the same objective, the minimization of cost during the recovery period, but differs from the network transition problem in the span of the transition period. Furthermore, the primary objective of the recovery problem from a network perspective is to return to the schedule as it was planned prior to the disruption as quickly as possible. Meeting this requirement may require a trade-off in the cost minimization objective. Capturing this trade-off can be enabled by allowing the algorithm to dynamically expand the recovery horizon as well as the set of vessels considered for the recovery operation. This approach is similar to that of Rezanova and Ryan (2006) where a train driver schedule recovery problem is considered and also uses the idea of heuristic problem expansion for the selection of aircraft in a recovery scenario proposed by Rosenberger et al. (2003). Finally, the recovery problem will typically be smaller than the network transition problem but on the other hand, the requirements to running time for the recovery problem will typically be more restrictive.

6.3 Solution Approach

As the network transition is fundamentally a rich form of the pickup and delivery problem (PDP), it is reasonable to expect this problem to share the properties of the PDP in terms of computational complexity. Given the size of realistic network transition problems and comparing this to the current capabilities of exact approaches to the PDP, a heuristic is believed to be the most viable approach to solving real-world problems.

We develop a large neighborhood search (LNS) heuristic to solve the network transition problem. In contrast to traditional local search heuristics that perform many small changes, e.g., 2-opt moves, to a solution to

search a larger portion of a defined neighborhood, the LNS heuristic is based on a series of neighborhoods resulting from larger changes to a solution but where only a small part of this neighborhood is searched. The LNS was originally proposed by Shaw (1998) and our implementation follows the same ideas. However, rather than extending the neighborhood incrementally when the search is stalled as suggested by Shaw (1998) and similar to the principle of variable neighborhood search (VNS) proposed by Hansen and Mladenović (2001, 2003) we use a fixed set of large neighborhoods. The large neighborhoods allow for the exploration of broader regions of the solution space thus mitigating some of the difficulties typically associated with traditional local search, e.g., escaping local minima. To allow for further diversification of the search, we embed the LNS in a simulated annealing framework.

Algorithm 1 lists the main structure of the adaptive LNS (ALNS) heuristic in pseudo-code. The listing corresponds to the flow of a single thread. In the implemented version of the ALNS, we run multiple threads in parallel and incorporate a common thread management and synchronization system. Threads collaborate on multiple levels. First, they exchange information about the global best solution, second, information about the performance of the individual neighborhoods is shared and finally warm start information is used when a thread stalls. Communication is always driven by the worker threads and thus no information is pushed or pulled by the thread management and synchronization system. For additional background on parallel strategies, Crainic and Toulouse (2003) provide a survey and discussion in the context of meta-heuristics.

Algorithm 1 Adaptive Large Neighborhood Search

```

1: Function ALNS( $s$ )
2:  $s_{localbest} = s$ 
3: repeat
4:   Select Destroy-Repair neighborhood  $i$ 
5:    $s' =$  Generate new solution from current solution  $s$  using  $i$ 
6:   if  $s'$  is accepted then
7:      $s = s'$ 
8:   if  $f(s') < f(s_{localbest})$  then
9:      $s_{localbest} = s'$ 
10:   Synchronize( $s_{localbest}$ )
11:   Update weight  $\omega_i$  of Destroy-Repair neighborhood  $i$ 
12:   if Reset criterion met then
13:     Reset  $s$  to GlobalBest
14: until Stop criterion met
15: return  $s_{best}$ 

```

The heuristic receives an initial solution as input and proceeds to initialize the local best solution $s_{localbest}$. In our implementation, each thread receives a different initial solution, the construction of which is discussed further in section 6.3.4. Next, Lines 4 and 5 encapsulate the core of the heuristic as this is where a new solution s' is constructed based on the current active solution s . This is achieved by first selecting a neighborhood defined as a pair of removal and insertion algorithms. The removal algorithm removes a subset of the currently serviced requests and the insertion algorithm subsequently reinserts these requests. We shall refer to this process as a neighborhood move using a corresponding neighborhood operator given by the selected pair of destroy and repair algorithms. At each iteration, we only generate a single solution from the neighborhood rather than searching the whole neighborhood. The destroy and repair principle is similar to the ruin and recreate heuristic proposed by Schrimpf et al. (2000) which also suggests to embed the search in a simulated annealing framework.

Lines 6–11 checks the new solution s' against an accept criterion, updates the local best solution if appropriate and synchronizes the found solution with the synchronization system which makes it accessible for other threads. Although we may search very large neighborhoods using the destroy-repair operators, we adopt a simulated annealing (SA) acceptance criterion rather than a strict descent criterion only accepting improving solutions as SA has shown to improve the search and solution quality. Weights determining the probability of selecting a particular neighborhood are updated in line 11.

Finally, Line 13 resets the current solution s to the best found solution across different threads if the search has stalled. We use a simple criterion and define a thread as stalled if it has not made any improvements to $s_{localbest}$ for a pre-defined number of iterations. The search terminates when the end temperature is reached which, in our implementation, translates into a certain number of iterations with appropriate selection of the cooling coefficient.

The ALNS framework is inspired by a similar approach proposed by Ropke and Pisinger (2006) and Pisinger and Ropke (2007) but differs in two key aspects. Where the framework of Ropke and Pisinger (2006) essentially operates with a single solution on which destroy-repair operations have been parallelized as well as a corresponding global temperature, we are running multiple threads of simulated annealing each with its own temperature and local best solution. We use periodic solution synchronization during the search and reset to the global best solution when a thread has stalled to allow the thread to proceed from this

solution. Secondly, the adaptive mechanism is based on destroy-repair neighborhood pairs in our implementation in contrast to the independent selection of destroy and repair methods adopted by Ropke and Pisinger (2006), Pisinger and Ropke (2007).

Although we use it here to solve the network transition problem, the ALNS framework is more general and can be used to solve many classes of optimization problems. In particular, the framework is well suited for tightly constrained problems where traditional methods relying on smaller neighborhoods tend to get trapped in local minima.

The following sections will describe the destroy and repair operators that have been implemented as well as the mechanisms for selecting these pairs and initializing the algorithm. It is worth noting that most of the described operators are applicable to other classes of pickup-and-delivery type problems and that some of the operators are indeed inspired by or adapted from the literature on vehicle routing and scheduling.

6.3.1 Destroy Algorithms

The destroy methods receive a (possibly partial) solution as input and proceeds to remove assigned requests from the routes according to a predefined procedure. Removed requests are inserted into a pool \mathcal{U} of unserved requests.

6.3.1.1 Random Removal

The simplest of the destroy algorithms is *random removal* which randomly selects and removes requests from an intermediate solution until a total of $q \leq |\mathcal{K} \setminus \mathcal{U}|$ requests have been removed. Each time a request k is removed from a route it must be verified that the resulting route remains feasible with respect to time windows. Enforcing the satisfaction of the triangle inequality and allowing waiting time, however, is enough to ensure feasibility. In our implementation, q is randomly selected within a parameterized interval $[qMin, qMax]$ the definition of which can greatly impact the performance of the algorithm. Section 6.4.2 will discuss this in further detail.

6.3.1.2 Related Removal

Simply removing requests at random, it may happen that the set of removed requests are not sufficiently related in terms of the imposed

constraints (time, load, and location) and that these thus maintain their original positions at a subsequent re-insertion. To mitigate this issue we define the related removal algorithm which removes visits based on a measure of relatedness. Initially proposed by Shaw (1998), this selection strategy tries to identify requests that are somehow related measured in terms of the constraints imposed on the assignment of requests to vessels. The idea is that requests that are related are expected to be more easily reassignable into objective improving positions.

We defined the relatedness $R(i, j)$ of two requests $i, j \in \mathcal{K}$ based on a weighted sum of their generalized distance from each other in terms of location, time window and volume (load). Introducing the weights α , β , and γ for the location, time window and load terms respectively, the relatedness measure is defined by

$$R(i, j) = \alpha(d_{l_P^i, l_P^j} + d_{l_D^i, l_D^j}) + \beta(|T_{l_P^i} - T_{l_P^j}| + |T_{l_D^i} - T_{l_D^j}|) + \gamma|d^i - d^j|. \quad (6.1)$$

The first term defines the geographical distance where $d_{l_P^i, l_P^j}$ denotes the distance between pickup locations l_P^i and l_P^j for i and j respectively and $d_{l_D^i, l_D^j}$ similarly for the delivery locations. Time window distance is defined as the absolute difference between actual service time with $T_{l_P^i}$ denoting service start of the pickup of i and $T_{l_D^i}$ similarly for the delivery. With this definition, the time window relates to the actual position in the route. Finally, the last term of (6.1) measures the absolute difference between the volumes of the two requests.

With the above definition, the lower $R(i, j)$ the “closer” and thus more related the two requests i and j are. Equation (6.1) is a slightly simplified version of the definition of relatedness in Ropke and Pisinger (2006) (and is inverse to that of Shaw (1998)).

The related removal algorithm is described in Algorithm 2. The algorithm takes as input a solution s , the maximum number of requests to remove q and a probability parameter p . At each step, the selection of the next related request is randomized with the degree of randomness controlled by the $p \in \mathbb{R}_+$. For $p = \infty$ the next request selected is the one with the highest degree of relation (lowest $R(i, j)$) to the currently selected seed request. As p decreases, the selection becomes increasingly randomized.

Algorithm 2 Related Removal

Function RelatedRemoval(s, q, p)
 r_s = Randomly select seed request from assigned requests K_s in the solution s
Initialize: $U = \emptyset$
while $|U| < q$ and $K_s \neq \emptyset$ **do**
 k = select request at random from U
 $G = K_s \setminus U$
 Sort G in order of increasing value of $R(\cdot, k)$
 y = select number randomly from $[0, 1)$
 $U = U \cup G[y^p | G|]$
Remove requests in U from s

6.3.1.3 Subsequence Removal

The principle behind subsequence removal is to remove a series of requests related in time. A seed request k_s is selected and the serving route $v_s(k_s)$ is identified. Based on the seed request and a time span δT , all requests serviced on route $v_s(k_s)$ between the service time $service(k_s)$ of k_s and $t_{end} = service(k_s) + \delta T$ are removed from the route and inserted into the set of unserved requests \mathcal{U} . Next, $q - 1 \geq 1$ routes are selected from $\mathcal{V}_T \setminus v_s$ at random and requests $k \in \mathcal{K}_v$ currently assigned to one of the selected routes v are removed if $service(k_s) \leq service(k) \leq t_{end}$. The principle has been illustrated in Figure 6.2. The rationale for this method of removing requests from a solution is that it frees sub-sections (both in time and geography) of two or more routes thus potentially allowing for the swapping of requests among these subsections. In the network transition problem requests are to be picked up and delivered at a limited set of geographical locations (ports). In this context, the subsequence removal allows for the potential swapping of visits to these locations (port calls) between two or more routes through the re-assignment of several requests. This neighborhood move may otherwise be difficult to achieve if e.g., requests are removed at random and multiple requests are picked up or delivered at a certain port location. Algorithm 3 outlines the flow of the subsequence removal algorithm.

6.3.2 Repair Algorithms

Once requests have been removed from an intermediate solution, the resulting set of unassigned requests \mathcal{U} is again reinserted using a series of simple insertion heuristics. Infeasible solutions are allowed during

Algorithm 3 Subsequence Removal

Function SubsequenceRemoval($s, q, \delta T$)
 k_s = Randomly select seed request from assigned requests in the solution s
Initialize: $\mathcal{U} = \mathcal{U} \cup k_s$
 $t_{end} = \text{ServiceStartT}(r_s) + \delta T$
Randomly select $q - 1$ routes from \mathcal{V}_T and insert into V'
for all $v \in V'$ **do**
 $k = \text{First request in } v \text{ with } \text{ServiceEndT}(k) \geq \text{ServiceStartT}(k_s)$
 while $t < T_{end}$ **do**
 remove k from route v
 $k = \text{Sucessor}(k)$
 $t = \text{ServiceStartT}(k)$

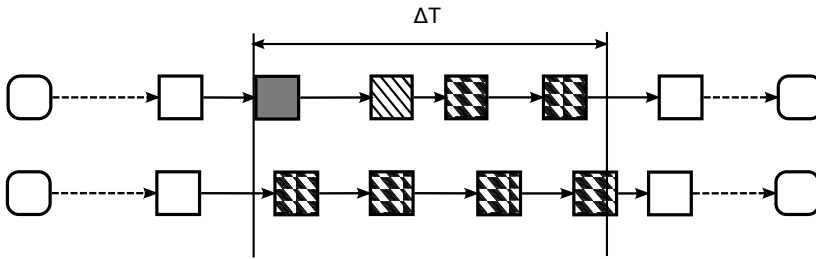


Figure 6.2: Subsequence removal selects a seed request (gray) and removes requests ΔT time intervals forward in the associated route and q randomly selected routes.

the search and we generally do not require that insertion heuristics are guaranteed to insert all requests in \mathcal{U} at each destroy-repair iteration. Although all of the insertion heuristics described below can be used to construct new solutions from scratch they will primarily be used in connection with partial solutions to insert requests that are not assigned to a route.

6.3.2.1 Greedy Insertion

One of the simplest insertion algorithms is the parallel greedy insertion heuristic. At each iteration we insert the request that increases the objective value the least. More specifically, let Δc_k^* denote the increase in the objective value by inserting request $k \in \mathcal{U} \subseteq \mathcal{K}$ in the best route at the best position. We set $\Delta c_k^* = \infty$ if no feasible insertion exist. At each iteration we insert $k = \arg \min_{k \in \mathcal{U}} (\Delta c_k^*)$ until all requests have been inserted or $\Delta c_k^* = \infty$ for all unassigned requests. We say that this heuristic is parallel, as opposed to sequential, because at each iteration we insert a request in the best route $v \in \mathcal{V}_T$ at the best position thus building routes in parallel.

The main benefit of the greedy insertion heuristic compared to more advanced insertion heuristics is the low computational complexity. Each insertion changes exactly one route and we only need to re-calculate the insertion cost of requests for which the corresponding route of best insertion has been changed. Also, given a complete distance matrix satisfying the triangle inequality, the algorithm performs at most $|\mathcal{U}|$ iterations since at each iteration we either insert a request or determine that the remaining requests are infeasible in the current solution.

6.3.2.2 Randomized Greedy Insertion

A randomized version of the greedy insertion heuristic randomly selects a request from the set of unassigned requests \mathcal{U} , determines the best insertion position among the available routes and inserts the request into this position if a valid position exists. This process continues until $\mathcal{U} = \emptyset$ or no more requests can validly be inserted in a route in \mathcal{V}_T . The main advantage of the randomized greedy insertion over the deterministic version is that we only need to calculate the insertion costs for each request once (again assuming the triangle inequality holds). Thus, the randomized greedy insertion heuristic is very fast. The randomized greedy insertion heuristic is also employed to construct the initial solution for each thread in PALNS.

6.3.2.3 Regret Insertion

One problem with the greedy insertion algorithm is its tendency to perform short-sighted decisions. When inserting a request, the greedy heuristic only evaluates the best position for all the unassigned requests and inserts the best among those. However, we will typically face situations where delaying the insertion of a request because it is not currently the best, will result in a later insertion cost which is significantly higher since alternative insertion positions may be much less attractive.

The regret insertion algorithm tries to mitigate this shortcoming by incorporating a look-ahead mechanism into the greedy insertion algorithm. The main idea is to insert the request that has the worst second best insertion cost relative to the cost of the best insertion of that request. Let Δc_k^1 denote the change in the objective value by inserting request $k \in \mathcal{K}$ at the best position in the best (cheapest) route and Δc_k^2 similarly denote the change in the objective by inserting in the second best route. At each iteration, the regret insertion heuristic then selects for insertion the request that satisfies the following:

$$k = \arg \max_{k \in \mathcal{U}} (\Delta c_k^2 - \Delta c_k^1) . \quad (6.2)$$

The regret principle outlined above was originally proposed by Tillman and Cain (1972) in connection with an algorithm for solving a capacity constrained routing problem with multiple depots. Martello and Toth (1981) later used the same principle to solve a generalized assignment problem. The regret measure can naturally be generalized to include the q subsequent route alternatives by summing the objective differences between the best and i 'th best alternatives, $i = 2, \dots, q$ as proposed by Potvin and Rousseau (1993) (using $q = |\mathcal{V}|$). The resulting q -regret selection criterion then becomes:

$$k = \arg \max_{k \in \mathcal{U}} \left(\sum_{i=2}^q (\Delta c_k^i - \Delta c_k^1) \right) \quad (6.3)$$

With this definition, the greedy insertion algorithm can be considered a 1-regret algorithm.

6.3.3 Destroy-Repair Pair Selection

Selection of a destroy-repair pair from the set η of neighborhood pairs is based on a roulette-wheel principle with weights assigned to the individ-

ual neighborhood pairs based on their historical performance. Letting ω_i denote the weight of a destroy-repair neighborhood pair $i \in \eta$, a neighborhood pair i is selected with a probability given by

$$p_i = \frac{\omega_i}{\sum_{j \in \eta} \omega_j} . \quad (6.4)$$

The weights of the individual destroy-repair neighborhood pairs $i \in \eta$ are updated according to a simple weighted moving average. Initially, weights are assigned the same value ω_i^0 , $i \in \eta$. At iteration n the weight of the selected destroy-repair pair is updated such that $\omega_i^n = \lambda \omega_i^{n-1} + (1 - \lambda)\sigma$. We refer to $\lambda \in [0, 1]$ as the dampening factor which controls the rate at which the weight of a destroy-repair pair i responds to the current performance. Large values of λ (close to 1) will result in a slow rate of change in the weights (high dampening) with $\lambda = 1$ yielding a fully random selection scheme. σ is a score measuring the success of a pair represented by three discrete cases; contributing with a new global best solution (highest score), contributing with a new local best solution, and finally no change to the best solutions but contributing with an accepted diversifying solution (lowest score). Both λ and the three scores are parameterized in the ALNS framework.

6.3.4 Initial Solutions

As previously mentioned, initial solutions are generated using the randomized greedy insertion algorithm. Although the quality of solutions produced from scratch using this algorithm is relatively low, it does provide a high degree of variation when constructing multiple initial solutions. When running the algorithm on multiple threads, this allows us to initially search different regions of the solutions space similar to the ideas behind multi-start heuristics, see e.g. Martí (2003) and Bräysy et al. (2004). Furthermore, experimental work suggests that the destroy-repair operations relatively quickly improve the initial solutions and that the ALNS algorithm is relatively robust with respect to poor starting solutions.

6.4 Computational Experiments

Computational experiments are executed on a computer with two quad-core Intel Xeon processors running at 2.66 GHz and with a total of 16

GB of system memory. The ALNS is implemented in C# .Net 3.5 with optimizations enabled. All reported times are wall clock time given in seconds.

6.4.1 Data

Although the development of the network transition problem is based on a real-world case, computational experiments are conducted on synthetic data generated by software specifically written for this purpose. This provides better options for controlling the properties of the problem instances as well as for generating a larger set of instances on which the ALNS is tested. Finally, the data generator allows for the construction of instances that are larger than those found in current real-world applications thus facilitating the testing of the scalability of the algorithm.

The synthetic data are divided into a series of classes covering problem sizes from small instances with 50-90 to large instances with up to 400 commodities. Within each class, several instances are created by adjusting a number of parameters defining size and characteristics.

A pre-defined set of ports is generated with locations distributed in an Euclidian plane with a pre-determined size. Port distribution falls in two categories. In the first category, locations are uniformly distributed within the plane constraints. The second category seeks to reflect the topology of the real-world case where a subset of the ports defined as hub ports are distributed in a smaller region in one of the extremes (corners) of the full problem plane. Remaining ports are distributed uniformly on the problem plane excluding the region containing the hub ports. The number of ports including possible hubs range between 10 and 30 in the instances generated.

Two sets of vessels are generated; the fixed-schedule vessels \mathcal{V}_F and the transitioning vessels \mathcal{V}_T . For each vessel $v \in \mathcal{V}_T \cup \mathcal{V}_F$ we generate a capacity, a cruising speed, a cost per nautical mile, and a canal cost all of which are drawn randomly from pre-defined intervals. The new schedule of each transitioning vessel is generated by randomly selecting ports and sequencing these together to form a cyclic schedule. The timing of the schedule is determined by selecting a time for the first port in the sequence and calculating the remaining call times based on distances, cruise speed, and average port time. The old schedule is represented by a single call to a port selected at random with a call time selected from the first part of the transition period. For the fixed schedule vessels \mathcal{V}_F the schedules are generated similarly to the vessels in \mathcal{V}_T but such that the cyclic schedule extends throughout the transition period. The length

of the transition period is either 7 or 14 days.

Commodities are generated by randomly picking an origin and a destination port as well as a time of availability which falls within the transition period $[T_{start}, T_{end}]$. Furthermore, we define a maximum allowed delay which together with the availability time defines the time window of the pickup. The overall freight flow (distribution of origin and destination) is determined by the port distribution category. For the clustered hub category, freight flows will be either hub \leftrightarrow non-hub or non-hub \rightarrow non-hub with the distribution determined by a parameter. For instances with hubs, approximately 90% of the flow will be hub \leftrightarrow non-hub.

The volume of each commodity is selected randomly from an interval with a fixed minimum and a maximum corresponding to a percentage of the capacity of the smallest vessel $v \in \mathcal{V}_T$. Revenue is determined randomly and is proportional to the volume of the commodity. We use a measure of time-volume given as the total time-volume of the transitioning fleet $\sum_{v \in \mathcal{V}_F} cap_v \cdot (T_{end} - T_{start})$ to control the total freight volume level. Each commodity consumes $T(l_P, l_D) \cdot d^k \cdot \gamma$ time-volume units where γ is referred to as the detour factor and represents a measure of the estimated excess time relative to $T(l_P, l_D)$ is required between pickup and delivery. No more commodities are generated when the total time-volume of the transitioning fleet is consumed. The detour factor γ controls the freight volume level relative to the fleet capacity.

6.4.2 Parameter Tuning

As there are several parameters for the ALNS algorithm as well as the individual neighborhood methods, parameter tuning is performed on a set of ten problem instances with a limited size. Tuning instances contain between 61 and 135 requests and 4-8 transitioning vessels. Each instance seeks to reflect a specific primary property and together the tuning instances represent a diverse set of problems in an attempt to obtain robust parameter settings. The primary properties fall in the following categories; randomly distributed requests, clustered random requests, high/low capacity to demand ratio, high/low commodity volume to vessel capacity ratio, and short/wide time windows. Instance properties are summarized in Table 6.1.

Random				Clustered			
Instance	$ \mathcal{V}_T $	$ \mathcal{V}_F $	$ \mathcal{K} $	Instance	$ \mathcal{V}_T $	$ \mathcal{V}_F $	$ \mathcal{K} $
T01	4	1	85	T06	4	1	61
T01	4	1	85	T07	4	2	61
T01	5	2	85	T08	5	2	88
T01	7	2	112	T09	7	2	135
T01	8	3	78	T10	8	3	75

Table 6.1: Overview of tuning instances.

6.4.2.1 Neighborhood Size

Two parameters `destroyMinPct` and `destroyMaxPct` control the size of the individual destroy-repair neighborhoods through restrictions on the minimum and maximum percentage of the requests that are removed and re-inserted in each iteration. The actual number of requests removed is selected uniformly at random from the interval $[\text{destroyMinPct}; \text{destroyMaxPct}]$ in each iteration. The values of these two parameters have a significant impact on the performance of the algorithm. Setting the upper limit too small can result in the algorithm becoming unable to escape local minima. On the other hand, setting the lower limit too high, means that a large portion of the solution is rebuilt at each iteration and solution quality then becomes much more dependent on the quality of the insertion heuristics. These are generally designed to be fast and do not produce high-quality solutions (as discussed in section 6.4.3) thus leading to little benefit from a significant amount of computational work.

Table 6.2 details the average objective values relative to the lowest value obtained by solving the tuning instances several times. For all but the first case (`destroyMaxPct` = 0.1) we set `destroyMinPct` = 0.1. The figures suggest that selecting an upper limit of 10% yields the best results although only 40% and 50% can be said to deviate significantly. For the largest instances, removing and re-inserting a large number of requests becomes inefficient and we additionally impose an absolute upper limit of 40 requests removed at each iteration.

Max pct	10%	20%	30%	40%	50%
Index	1.000	1.009	1.006	1.014	1.019

Table 6.2: Effect of the size of the neighborhood (measured in terms of the percentage of removed requests).

6.4.2.2 Start and End Temperature of the Simulated Annealing

Setting the start (and end) temperatures of the annealing acceptance criterion is crucial for the quality of the solutions produced. Setting the start temperature too high may result in erratic behavior of the algorithm with the frequent acceptance of possibly very poor-quality solutions. A significant amount of computational work will thus be spent without improving the solution quality. On the other hand, setting a too low start temperature essentially turns the ALNS into a strict descent algorithm with the risk of getting trapped in a local minimum. The challenge is that the appropriate start and end temperatures are dependent on the specific problem instance. To mitigate this issue, we have introduced a pre-burner which essentially runs a limited number of descent iterations on the problem collecting statistics on the development of the objective function value. Based on predefined acceptance criteria at the beginning and end of the simulated annealing process we can then set a start and end temperature appropriate for the order of magnitude of the objective function value of a specific problem instance.

6.4.3 Performance of the Insertion Heuristics

The performance of the individual insertion algorithms has been evaluated by using these to construct initial solutions from scratch. Both the greedy and the regret heuristic are deterministic and are thus only run once on each tuning instance. The randomized greedy insertion heuristic has been run several times and the numbers reported in Table 6.3 are averages over these runs. *Time* indicates the total time in seconds (s) spent constructing initial solutions for all of the ten tuning instances. The row *Avg. unassigned* indicates the average proportion (in percent) of requests that are not assigned in the initial solution. Finally, *Avg. gap* provides the average over the tuning instances for the gap between the objective value of the initial solution and the best known solution relative to the best known solution.

None of the insertion heuristics manage to consistently construct initial solutions in which all requests are served. Overall, the regret-2 algorithm performs best, producing initial solutions serving the largest proportion of requests on average. This was to be expected as the regret-2 algorithm incorporates more information than the other two insertion heuristics. However, improved quality comes at a cost in terms of increases in the computational time required. What is surprising, however, is that for

certain instances the randomized greedy insertion outperforms regret insertion in terms of the proportion of requests that are serviced in the initial solutions. This behaviour may possibly be attributed to lucky sequencing of the insertions.

	Random greedy	Greedy	Regret
Avg. gap (%)	64.5	63.6	58.9
Avg. unassigned (%)	19.1	18.6	16.8
Time (s)	0.308	0.673	0.849

Table 6.3: Comparison of the performance of insertion heuristics when used to construct initial solutions. Figures are based on running each heuristic on each of the tuning instances. Time is total time used for all instances.

The ALNS heuristic is particularly well suited to solve highly constrained problems where it can often be difficult to move between feasible solutions. The network transition problem is in fact a problem where simply obtaining feasible solutions is difficult. This is evidenced by the statistics provided in Table 6.3 and supports the need for allowing infeasible solutions during the search or adopt a revenue/penalty perspective which allows requests to remain unserved.

6.4.4 Effect of the Cooperative Multi-threading

Intuitively, we expect solution quality (measured in terms of the objective value) to increase with the number of threads given the same amount execution time since more computational work is performed and the individual threads cooperate. Figure 6.3 illustrates and supports this expectation by plotting the development of the objective value as a function of time when using one, two, four, and eight threads respectively. The curves plotted represent the average over ten runs of a single instance for each of the thread number settings.

When using only a few threads, we observe longer periods where the objective remains unchanged (horizontal parts of the curves). Also worth noting is that with an increase in the number of threads we also see an increase in the overhead incurred by thread synchronization. In each of the four cases illustrated in Figure 6.3, each thread performs a fixed number of iterations but the total running time increases from 293 seconds on average for the one-threaded case to 357 seconds for the eight-threaded case. Additional computational overhead is also observed when consid-

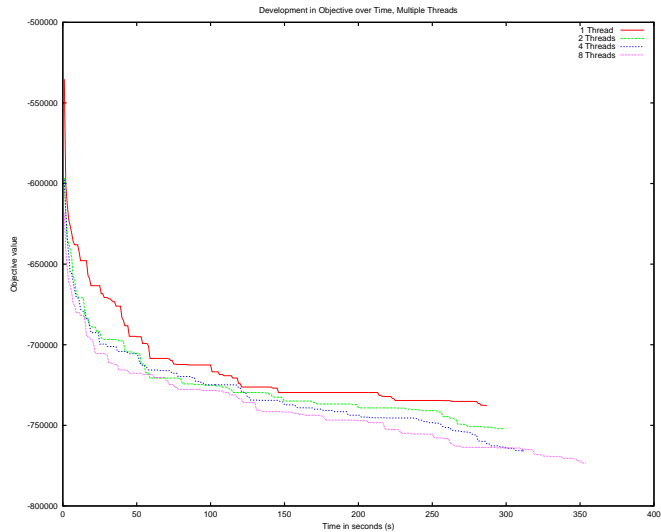


Figure 6.3: Development of the objective over time based on one, two, four, and eight threads respectively. Instance: T07

ering the multi-thread speed-up factor calculated as the time required to reach a certain objective value using one thread, relative to the time required using n threads. Speed-up factors for one, two, four, and eight threads are provided in Table 6.4. The relatively low speed-up factors observed may be explained by the solution synchronization mechanism potentially resulting in the re-discovery of solutions by multiple threads without a corresponding improvement in the global best solution.

	1 Thread	2 Threads	4 Threads	8 Threads
Time (s)	100	56	54	38
Speed-up	1	1.79	1.85	2.63

Table 6.4: Speed-up factor ($Time(1)/Time(nThreads)$) with the objective value obtained using one thread at 100s as target.

Adopting a fixed-work perspective rather than fixed-time as above, the picture expectedly changes somewhat. Performing a series of experiments with between one and eight threads where the total number of iterations is fixed, results are obtained that support the conclusions of Cordeau and Laporte (2003) that performing more iterations on fewer threads is better than performing fewer iterations on more threads. This behaviour is explained by the fact that performing only a few iterations

on many threads does not allow any of the threads to search deep enough into the solution space. Furthermore, as all threads are started with initial solutions that are expected to be equally far from a local optimum, the cooperative mechanism is only of limited value in cases where none of the threads reach a local minimum.

6.4.5 Scalability

To evaluate the scalability of the algorithm in terms of computational time a series of experiments are conducted for a set of 43 test instances divided into three classes; small (50–90 commodities), medium (~ 150 –200 commodities), and large (~ 250 –400 commodities). Note that the literature on the pickup and delivery problem usually report problem sizes in terms of the number of requests where our definition of a single commodity corresponds to two requests (a pickup and a delivery). Thus, using this measure, the largest instances correspond to 800 (pickup or delivery) requests.

All instances are solved using the best settings found in the previous analysis. The number of iterations performed for each instance is fixed within each problem class and selected such that the algorithm terminates when improvements in the objective value start to level off.

Tables 6.5–6.7 provide an overview of the results of solving the 43 test instances using the ALNS algorithm. For each instance we summarize the problem size in terms of the number of requests (nReq) and the number of transitioning vessels (nVes). Each instance has been solved five times and we report the average wall clock time in seconds required to solve the instance (Avg time), the average fraction of the total time spent to obtain the best solution within each run (Avg T to best), and the average solution quality relative to the “best known” solution (Avg obj qual). “Best known” solutions are obtained by running the ALNS algorithm for a large number of iterations.

For the largest test instances average solution time is over four hours. However, this time must be evaluated in the context of the application and the planning horizon on which the model is employed. Thus, for determining and evaluating transition schedules, four hours is an acceptable solution time given that the process is carried out weeks in advance of actual implementation. With regards to the time to the best solution for particularly the medium and large instances, these are all close to the total running time indicating that improvements to the objective were still being made when the algorithm was terminated. However, these improvements were generally quite small, in the order of 10^{-3} relative

Instance	nReq	nVes	Avg time	Avg T to best	Avg obj qual
S01	50	6	134	0.87	1.02
S02	50	8	113	0.93	1.07
S03	50	8	126	0.82	1.66
S04	50	8	145	0.82	1.16
S05	49	8	128	0.88	1.02
S06	50	8	130	0.93	1.04
S07	50	8	139	0.95	1.02
S08	50	8	141	0.92	1.06
S09	50	8	131	0.95	1.10
S10	70	5	429	0.67	1.19
S11	70	5	1001	0.93	1.02
S12	70	5	639	0.53	1.00
S13	70	4	926	0.22	1.03
S14	90	5	738	0.88	1.09
S15	89	5	511	0.75	1.20
S16	90	6	578	0.93	1.04
<i>Avg</i>			<i>376</i>	<i>0.81</i>	<i>1.11</i>

Table 6.5: Results for the Small-class test instances.

Instance	nReq	nVes	Avg time	Time to best	Avg obj qual
M01	200	10	4355	0.89	1.27
M02	195	10	4583	0.98	1.13
M03	187	10	2991	0.91	1.09
M04	200	10	2325	0.95	1.03
M05	149	10	1916	0.97	1.06
M06	200	10	2730	0.98	1.08
M07	200	10	3656	0.96	1.16
M08	200	10	3163	0.98	1.06
M09	200	15	2571	0.96	1.11
M10	200	15	2676	0.97	1.10
M11	200	15	2009	0.97	1.08
M12	200	10	4069	0.97	1.06
M13	200	10	3821	0.96	1.06
M14	200	15	2491	0.97	1.05
M15	200	15	2015	0.98	1.15
M16	200	15	3240	0.94	1.05
M17	200	15	3167	0.96	1.09
<i>Avg</i>			<i>3046</i>	<i>0.96</i>	<i>1.10</i>

Table 6.6: Results for the Medium-class test instances.

Instance	nReq	nVes	Avg time	Time to best	Avg obj qual
L01	361	15	15417	0.96	1.07
L02	383	15	22388	0.88	1.22
L03	271	17	8583	0.97	1.04
L04	339	15	15978	0.96	1.08
L05	400	15	21488	0.89	1.08
L06	371	15	17328	0.96	1.08
L07	371	15	21399	0.94	1.09
L08	282	17	9803	0.95	1.04
L09	316	15	13429	0.98	1.16
L10	397	15	20179	0.97	1.07
<i>Avg</i>			<i>16599</i>	<i>0.95</i>	<i>1.09</i>

Table 6.7: Results for the Large-class test instances.

to the objective value, justifying the termination of the algorithm. Figure 6.4 provides an example of this behaviour for one of the medium class instances.

Overall, solutions are within 11% of the best known solutions although with some significant deviations for certain cases, e.g., S03, M01, and L02. In general, however, the ALNS is robust towards different instance characteristics with small deviations in objective value between different runs.

6.4.6 A Real-World Case

Where synthetic data provide a high degree of control over various properties of the individual instances, such data tend to exhibit more randomness than real-world data, in particular in the demand flow patterns. One of the effects of this randomness is the relatively low capacity utilizations that can be achieved due to a complex interplay between locations and time windows of individual requests. This also means that one of the central key performance indicators for a liner service network, the utilization of vessel capacity, can not be reliably evaluated for the synthetic instances. In real-world instances, on the other hand, there is an natural correlation between the demand flow patterns and the routes operated in the service network.

To evaluate the performance of the ALNS in a setting more close to a real-world scenario, an instance has been constructed based on actual data from a feeder shipping company. This case contains 11 vessels and

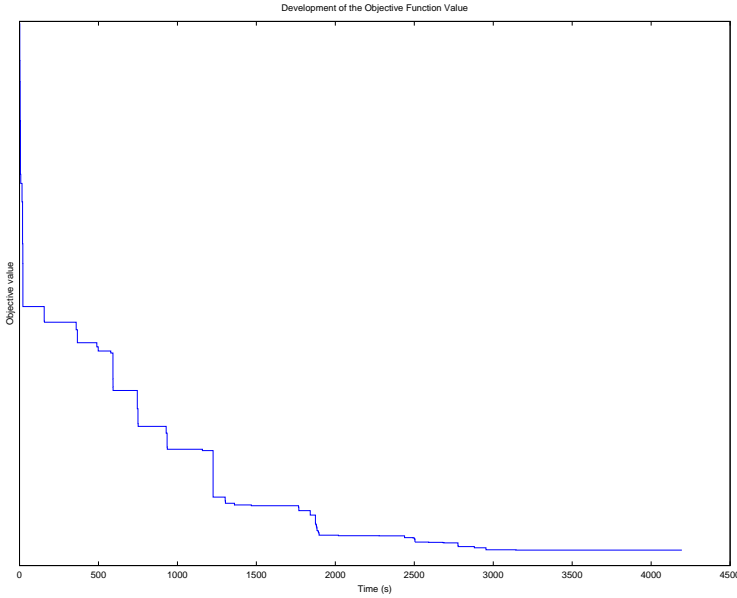


Figure 6.4: Example of the development of the objective function value.

218 requests and is constructed by taking the actual schedule of these 11 vessels and then deleting the planned port calls during the transition period. Essentially, this corresponds to performing a transition from and to the same schedule and allows us to compare the ALNS proposed schedule with the schedule actually operated.

The instance is solved in approximately two and a half hours resulting in a solution with an average capacity utilization across all vessels of 63% relative to the nominal volume capacity of the vessels with individual vessels seeing utilizations between 51%–74%. While these may initially not seem like a very high levels of utilization of the vessel capacity, such a measure is naturally dependent on the available freight. Furthermore, a number of operational constraints such as stowage are not handled in the current algorithm. This means that the nominal capacity of a vessel only provides an approximate image of the actual capacity which may be lower, thus leading to lower figures when used as the basis for utilization calculations. Further analysis shows that capacity utilization per journey leg is highest on the legs going between hub ports and non-hub ports while showing lower values on the inter non-hub legs. This corresponds to the behaviour in the real-world operations.

Given the difficulties of comparing capacity utilization between the solutions obtained using the ALNS heuristic and the original schedule a more reliable performance indicator is the total sailing distance required to cover the available freight. With respect to this measure, the ALNS solution offers a reduction of 6.9% in the total distance sailed compared to the original schedule. Translated into monetary value this corresponds to potential savings of more than \$50,000 over the transition period given late 2009 fuel prices.

Continuing the analysis of the potential for savings, we modify the above instance by removing one of the poorest performing vessels thereby reducing the system capacity. The result is a further decrease in the total distance sailed (15% compared to the original schedule). Most interesting however, is that all demand is still met but with one less vessel resulting in potential time charter savings of around \$60,000 in addition to more than \$100,000 in fuel savings over the transition period relative to the original schedule.

Finally, since solutions are obtained in much less time that required for manual planning, the use of the ALNS adds significant value in a planning situation where different scenarios are evaluated. This is true even in situations where only minor improvements to the key performance indicators are achieved.

Looking further into the solutions, it is observed that although demand is derived from an operated cyclic schedule, the transition schedules produced by the ALNS algorithm contain little repetition in the schedules of the individual vessels. Schedules also tend to cover ports from multiple regions which is in contrast to the schedule operated by the feeder shipping company where vessels only visit a limited number of ports typically within only a few regions. Although such a schedule might not be desirable for a master schedule, it is not considered a problem in relation to the current problem context; the transition from one service network (and schedule) to another.

6.5 Conclusions and Further Work

In this paper we have presented a new problem, the network transition problem, which attempts to bridge the gap between designing a new service network and actually implementing this new design in the operational network. The problem shares features with the classic pickup and delivery problem with time windows but extends this problem with several additional features.

A parallel cooperative adaptive large neighborhood search (ALNS) heuristic has been developed to solve the problem. We proposed a series of neighborhood operators in the form of removal and insertion algorithms. Some of the operators were adapted from the vehicle routing literature while others were developed specifically for the network transition problem. We have conducted an analysis of the ALNS and found it to be robust relative to different settings for the parameters although it is also clear that selecting a good set of neighborhood algorithms capable of performing both diversification and intensification in the local search is important for the quality of the solutions produced.

Computational experiments on a set of randomly generated test instances showed that the ALNS is capable of solving realistically sized problems in reasonable time. Furthermore, these experiments indicated that the ALNS is robust against different instance characteristics. In a more realistic scenario defined by a real-world data set, the ALNS produced schedules of a quality comparable to those operated by the data owner but using less planning time than required by manual planning. Additionally, the ALNS produced schedule offered a savings potential of more than \$50,000 over the transition period with the same fleet and more than \$160,000 when allowing the fleet to be reduced by a single vessel.

One very interesting perspective of this work is the opportunity for the liner service provider to move towards a dynamic planning process with a rolling time horizon. Basically, this would correspond to performing frequent network transitions but instead of transitioning to a new service network, the requirement would be to return to the schedule defined by a master plan. The main motivation for the implementation of such a revised planning process would be the observation that even in the current planning process, frequent but minor schedule updates are performed to adapt to changes in the operating environment such as demand volumes or resource availability (schedule disruptions). However, these minor updates generally do not take a sufficiently broad perspective on the consequences for the whole network and may often prove costly to the liner service provider.

Acknowledgement

This work was partially supported by The Danish Maritime Fund. This support is gratefully acknowledged.

Bibliography

- M. W. Andersen, C. Barnhart, and O. B. G. Madsen. Service network design in a liner container feeder application. Institute of Transport, Technical University of Denmark, Working paper, 2009.
- M. Ball, C. Barnhart, G. Nemhauser, and A. Odoni. Air transportation: Irregular operations and control. In C. Barnhart and G. Laporte, editors, *Transportation*, volume 14 of *Handbooks in Operations Research and Management Science*, chapter 1. North-Holland Publishing Company, 2007.
- G. Berbeglia, J.-F. Cordeau, I. Gribkovskaia, and G. Laporte. Static pickup and delivery problems: A classification scheme and survey. *TOP*, 15(1):1–31, 2007.
- O. Bräysy, G. Hasle, and W. Dullaert. A multi-start local search algorithm for the vehicle routing problem with time windows. *European Journal of Operational Research*, 159(3):586–605, 2004. ISSN 03772217.
- M. Christiansen, K. Fagerholt, and D. Ronen. Ship routing and scheduling: Status and perspectives. *Transportation Science*, 38(1):1–18, 2004. ISSN 00411655.
- M. Christiansen, K. Fagerholt, B. Nygreen, and D. Ronen. Maritime transportation. In C. Barnhart and G. Laporte, editors, *Transportation*, volume 14 of *Handbooks in Operations Research and Management Science*, chapter 4. North-Holland Publishing Company, 2007.
- J.-F. Cordeau and G. Laporte. A tabu search heuristic for the static multi-vehicle dial-a-ride problem. *Transportation Research Part B: Methodological*, 37(6):579–594, 2003. ISSN 0191-2615.
- C. E. Cortés, M. Matamala, and C. Contardo. The pickup and delivery problem with transfers: Formulation and a branch-and-cut solution method. *European Journal of Operational Research*, 200:711–724, 2010.
- T. Crainic and M. Toulouse. Parallel strategies for meta-heuristics. In *Handbook of Metaheuristics*, volume 57 of *International Series in Operations Research & Management Science*, chapter 17, pages 475–513. Springer New York, 2003.
- T. G. Crainic. Service network design in freight transportation. *European Journal of Operational Research*, 122:272–288, 2000.

- P. Hansen and N. Mladenović. Variable neighborhood search: Principles and applications. *European Journal of Operational Research*, 130:449–467, 2001.
- P. Hansen and N. Mladenović. Variable neighborhood search. In *Handbook of Metaheuristics*, International Series in Operations Research & Management Science, chapter 6, pages 145–184. Springer New York, 2003.
- S. Martello and P. Toth. An algorithm for the generalized assignment problem. In *Proceedings for the Ninth IFORS International Conference. Operational Research '81.*, 1981.
- R. Martí. Multi-start methods. In *Handbook of Metaheuristics*, volume 57 of *International Series in Operations Research & Management Science*, chapter 12, pages 355–368. Springer New York, 2003.
- S. Mitrović-Minić and G. Laporte. The pickup and delivery problem with time windows and transshipment. *INFOR*, 44(3):217–227, 2006.
- S. N. Parragh, K. F. Doerner, and R. F. Hartl. A survey on pickup and delivery problems - part i. *Journal für Betriebswirtschaft*, 58(1):21–51, 2008a.
- S. N. Parragh, K. F. Doerner, and R. F. Hartl. A survey on pickup and delivery problems - part ii. *Journal für Betriebswirtschaft*, 58(2): 81–117, 2008b.
- D. Pisinger and S. Ropke. A general heuristic for vehicle routing problems. *Computers & Operations Research*, 34:2403–2435, 2007.
- J.-Y. Potvin and J.-M. Rousseau. A parallel route building algorithm for the vehicle routing and scheduling problem with time windows. *European Journal of Operational Research*, 66:331–340, 1993.
- N. Rezanova and D. M. Ryan. The train driver recovery problem - a set partitioning based model and solution method. Technical Report IMM-Technical Report-2006-24, Technical University of Denmark, Informatics and Mathematical Modelling, 2006.
- S. Ropke and D. Pisinger. An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows. *Transportation Science*, 40(4):455–472, 2006.
- J. M. Rosenberger, E. L. Johnson, and G. L. Nemhauser. Rerouting aircraft for airline recovery. *Transportation Science*, 37(4):408–421, 2003.

- G. Schrimpf, J. Schneider, H. Stamm-Wilbrandt, and G. Dueck. Record breaking optimization results using the ruin and recreate principle. *Journal of Computational Physics*, 159:139–171, 2000.
- P. Shaw. Using constraint programming and local search methods to solve vehicle routing problems. In M. Maher and J.-F. Puget, editors, *Principles and Practise of Constraint Programming – CP98*, volume 1520 of *Lecture Notes in Computer Science*, pages 417–431. Springer Berlin, 1998.
- F. A. Tillman and T. M. Cain. An upperbound algorithm for the single and multiple terminal delivery problem. *Management Science*, 18(11): 664–682, 1972.
- P. Toth and D. Vigo, editors. *The Vehicle Routing Problem*. Discrete Mathematics and Applications. SIAM, 2002.

Comments on Papers

7.1 Paper 1: Service Network Design in a Liner Container Feeder Application

Paper 1, “Service Network Design in a Liner Container Feeder Application”, discussed a service network design problem in a maritime setting and proposed a new approach to problem decomposition by exploiting features specific to the case that inspired the work. A simple heuristic was used to generate route candidates in a sub-problem using a reduced cost estimate based on dual information from a master problem. The ability of the route generator to provide high quality routes will greatly impact the overall solution quality and is thus a candidate for further analysis. First steps in this analysis and comments on an approach to improve the route generator is provided in the following section.

7.1.1 Advanced Route Construction Heuristics

One of the key challenges in ensuring high quality solutions using the decomposition principle employed in paper 1 is the construction of good routes. The current approach is based on a very fast but simple heuristic building routes by simply selecting the most promising commodities

(measured in terms of contribution to reduced cost estimate) in a greedy fashion. It is clear however, that such an approach does not take into account the more broad consequences of the selection of a particular commodity, such as the set of alternative commodities that are rendered infeasible by such a decision. Ideally, new routes should be constructed through implicit enumeration of all candidates e.g., through the use of a resource constrained shortest path algorithm on an appropriately constructed network. This approach, however, is much too expensive and partially defeats the idea behind the decomposition.

An alternative to the expensive implicit enumeration is the improvement of the construction heuristic by extending it with some sense of global impact evaluation. Based on the discussions in paper 2 it would seem obvious that using a regret-based insertion heuristic (see page 117 for a description) would meet this requirement. However, complications are associated with measuring the cost of the “second best” insertion in the present setup as routes are built independently for each class of vessels. These complications may be overcome by constructing routes in parallel as is done in paper 2. Such a modification would have the added benefit of more closely reflecting the fact that all commodities should in general be covered. Parallel route building could then be further extended by a limited local search following a destroy-repair principle similar to that of the adaptive large neighborhood search algorithm described in paper 2, chapter 6. Even a less complex modification of the current heuristic implemented by performing a limited number of random removal and re-insertion operations during sequential route construction could lead to improvements of the overall solution quality.

7.2 Paper 2: Network Transition in a Liner Container Shipping Application

Paper 2, “Network Transition in a Liner Container Shipping Applications”, presented a parallel cooperative adaptive large neighborhood search algorithm (ALNS) to solve a new problem dubbed the *network transition* problem. The problem shares many features with the pickup and delivery problem with time windows but extends this problem with additional properties such as multiple alternative delivery locations. Due to the flexible and diverse nature of the ALNS and the network transition problem, relatively simple extensions or adaptations will enable the application of the framework to a series of closely related problems.

The following section will provide supplemental suggestions to possi-

ble enhancements of the ALNS algorithm and section 7.2.2 will discuss the ALNS in the context of continuous planning with a rolling horizon. Subsequently, section 7.2.3 will describe the challenges associated with adapting the network transition algorithm to solve the network recovery problem previously described in section 2.5.

7.2.1 Algorithmic Extensions and Enhancements

One of the issues that were occasionally observed with the ALNS when solving large highly constrained instances was periods during the search with a relatively large number of non-improving iterations followed by a short series of objective improving iterations before stalling again. Although embedded in a simulated annealing framework to facilitate diversification, it seems that the mechanisms provided are not always sufficient to quickly escape local minima. Even though the start and end temperatures of the simulated annealing are carefully set according to the scale of the objective, many problems may observe large differences between the objective value of the initial solutions and the objective value upon termination. If convergence of the objective value is not sufficiently smooth relative to the temperature reduction during the search, this behavior leads to an imbalanced acceptance rate. This means that too many non-improving solutions may be accepted after an initial strong drop in objective value leading to an erratic search. Similarly, when the objective value convergence has leveled out, the reduced temperature complicates the escape from the current local minimum.

Instead of relying exclusively on the simulated annealing to diversify the search it may be beneficial to incorporate a diversification mechanism which is triggered when the algorithm has reached a predefined criterion indicating that it has stalled. The idea is somewhat similar to using iterated local search (see e.g., Lourenço et al., 2003). However, rather than building an entirely new solution using the construction heuristics (which are known to produce quite poor quality solutions), it may be beneficial to use the existing destroy and repair neighborhood methods but with more aggressive settings than those used during normal neighbor creation. Some preliminary work has been done pursuing this idea, but no clear conclusions can be made at this point. It is, however, clear that particular care must be taken when defining the stalled criterion as well as the extent to which the active solution should be destroyed if the algorithm is not to degenerate into a multi-start heuristic. See e.g., Martí (2003) for an introduction to multi-start heuristics and Hansen and Mladenović (2003) for a comment regarding the potentially degenerate nature of such heuristics. Furthermore, the interplay between the diver-

sification method and the simulated annealing complicates the analysis as the temperature will impact the convergence characteristics following a diversification.

Another important aspect of the ALNS is the synchronization mechanism which differentiates the algorithm from a traditional multi-start local search. The synchronization mechanism is responsible for the sharing of solution information between the different threads of the ALNS. However, this information is used only in connection with reset operations in which the active solution of a thread is updated to the global best solution when the thread has not contributed with a new global best solution for a set number of iterations. Although neighborhood searches are randomized in the ALNS, it may be beneficial to augment the synchronization mechanism with a more extended type of memory which could then be used by threads to reduce the number of unpromising solutions searched. Crainic and Gendreau (2002) investigates several aspects related to the use of a central solution pool in a Tabu search heuristic, including which type of information to share and how this information should be used by the individual threads. Based on experiments with several strategies for maintaining the solution pool and exchanging information with the threads, the authors conclude that definition of these strategies is critical to the performance of the algorithm. Thus, it is not a trivial task to extend the synchronization mechanism. It remains clear, however, based on the work with the ALNS and supported by Crainic and Gendreau (2002) that even a simple synchronization of information about the current best known solution can significantly improve solution quality provided that synchronization intervals are selected to allow for a sufficient level of diversified search.

7.2.2 Using ALNS in Continuous Planning

In a world with perfect information (at least in the short term), schedules and service networks would be adapted to known demand and be updated dynamically and frequently as new information became available. Unfortunately, the information about demand that is currently available at the time when the service network must be designed and schedules published is far from perfect. However, even with this less than perfect information, it may be possible to design the service network following the current principle based on some average demand scenario but at the same time incorporate properties into this network making it more amenable to dynamic rescheduling. Some of these properties have previously been described in section 2.5.2 and include frequent route crossings of vessels with different characteristics, schedule slack and multiple vessels sup-

plying coverage of a particular region. In general, these are interpreted as properties of robust networks. Under these conditions, the network transition problem and associated solution approach would enable the partial implementation of dynamic re-scheduling. Subject to the constraint that the published schedule should be met within some allowed tolerance, the transition would be characterized as defining a temporary schedule for a subset of vessels with current vessel positions as the initial state and the currently published (and possibly operated) schedule as the end conditions. In this respect, the only difference from the network transition problem as it is formulated in chapter 6 is that the transition is not to a new service network with a new schedule but simply to the preexisting service network and schedule. This process could then be repeated on e.g., a weekly basis thus providing a continuous planning with a rolling horizon while only performing short term small modifications to the schedule.

It may be argued that a continuous planning process will defeat the previously described requirements to service reliability, schedule conformance, and evaluation of network wide consequences of changes (see section 2.1). However, when analyzed further, it turns out that the current practice of many liner carriers, particularly short sea, is to perform frequent ad hoc changes to the network to cater for changing operating conditions, mostly in the form of changes to demand (volume changes and spot freight). The main issue related to this practice is not necessarily the changes themselves but rather the fact that consequences are only evaluated on a very short term local scale making it almost impossible to fully capture the economical impact of such decisions. Employing the network transition solution framework would facilitate a more broad evaluation of changes even if the framework was only allowed a limited degree of freedom thus alleviating some of the concerns regarding economical impact assessment.

Taking the idea of dynamic scheduling even further, it is tempting to imagine a setup where the schedule is not expressed as a strict timetable but rather an overall service specification allowing some additional flexibility in the actual scheduling. Subject to the service specification, the schedule would then be dynamically updated as new information became available. Such a setup is very similar to the way that tramp shipping currently operates and will put very high requirements on forecast systems and coordination with customers to be viable in a liner shipping environment competing on service reliability and schedule conformance. Given the current state of planning and forecasting tools as well as research on the subject, such an endeavor is not believed to be feasible in the near future.

7.2.3 Adapting to the Network Recovery Problem

Section 2.5 briefly introduced the network recovery problem but did not address the question of how the process of recovery can be executed. As it turns out, the perspectives of the framework developed in Paper 2, chapter 6 extend even further than applications to network transition and continuous planning.

The basic idea remains the same as in the network transition problem; a subset of the fleet defines the vessels that are part of a recovery plan and initial and end conditions are defined by the current position of these vessels and their original schedule at some later time, respectively. The challenge is defining the set of vessels that can be used in the recovery plan and the time horizon that the recovery plan should cover. This latter aspect will restrict how much time is allowed for recovery and will be defined by both practical considerations, i.e., how fast can a recovery feasibly be achieved?, and by business restrictions, i.e., what is the longest period that will be allowed for the recovery? The longer the recovery period, the greater the risk of new disruptions occurring becomes. Furthermore, allowing a too long recovery period may mean that overall schedule compliance falls below a pre-defined acceptable level. On the other hand, a too short transition period can result in either an expensive recovery plan or no feasible plan at all.

In case of both the selection of recovery vessel set and recovery period, a dynamic expansion strategy can be adopted. The idea is to initialize the vessel set and period according to some selection rule and then dynamically expand the period and/or the vessel set if no satisfactory or feasible recovery plan is found. The challenge then becomes determining appropriate initial conditions and the rules according to which the subsequent expansions can occur. Although some literature deals with these problems in the context of e.g., airline recovery (see e.g., Clausen et al. (2010) for a recent survey and analysis) and railway crew recovery (Rezanova, 2009), specific initialization and selection rules should always be tailored to the specific business employing the algorithmic framework. However, an obvious choice for the definition of the initial recovery vessel set is to include all disrupted vessels. This set can then be extended by “nearby” vessels defined as some combination of temporal and geographical proximity possibly augmented with a measure of likeness in terms of both physical qualities as well as the regions serviced by vessel candidates. With respect to the length of the recovery period, a good rule for the initial duration can probably be determined and combined with a constraint on the maximum allowed duration. The algorithm can then switch between expanding first the vessel set a pre-defined number

of times and then as a secondary action, expand the recovery period

Conclusion

Liner based container shipping has been one of the primary enabling factors in the globalization of supply chains and plays a major role in modern global transportation of manufactured goods. Given this significant role, it remains surprising that a business area that has such a large impact on many levels of both the global economy as well as local economies, from globalized supply chains to improving product choice for the individual, has received such little attention from the operations research community. Particularly considering that from a research perspective, the problems involved in liner based container shipping are both highly challenging and diverse.

This thesis has sought to provide the first steps toward a revitalization of the research within liner based freight transportation by addressing one of the central planning processes faced by liner container carriers; the liner service network design problem. The contributions toward meeting this goal are three-fold. First, a unified description of the liner service network design problem has been provided, emphasizing the crucial connections to planning problems faced by carriers at different planning levels. Second, an integrated development of a series of models and solution methods for the service network design problem has been presented evaluating the merits of each of the models with respect to supporting the planning processes currently executed by carriers as well as opportunities for incorporating expected future trends affecting these processes.

Finally, a concrete model from this presentation has been analyzed further in the context of a real-world container feeder service network design problem and a solution method has been developed. The critical issue of ensuring business adoption and implementation of the service network design model has been addressed through the development of algorithms for the solution of a problem new to the literature, the network transition problem, thus providing an important tool in linking current and future service networks.

Two papers are included in this thesis addressing the liner container feeder service network design problem and the network transition problem respectively.

Inspired by a real-world case in a liner container feeder application, the first paper addressed a concrete service network design problem. Using problem specific characteristics allowed for development of a solution approach based on a new decomposition principle. Central to this approach was the idea of maintaining a route pool which was iteratively augmented using a heuristic to determine new route candidates in a route generating sub-problem. Additionally, rather than maintaining all the routes of the route pool in a master problem, the solution approach used a selection mechanism based on dual estimation to determine attractive routes. The dual estimates are also used by the route construction heuristic. Demand was modeled through the concept of commodity groups which integrates multiple commodities into a single variable representation. Dynamic generation of these commodity groups was achieved through a series of route dependent packing sub-problems such that a commodity group corresponded to a feasible route packing. Although the packing sub-problems depend on the individual routes of which there may be many, only a small number of routes are active at any point due to the route pool maintenance and selection criterion. Thus, the advantages associated with the rich set of problem dimensions that can be modeled using this decomposition approach outweigh the potential complexity of maintaining many route specific sub-problems. The model currently represents one of the most rich representations of the liner service network design problem.

Computational experiments on a series of instances based on real-world data showed the tractability of the solution approach. The developed algorithm was capable of solving the instances of a realistic size within a time frame appropriate for the tactical nature of the service network design problem. Several methods for managing and augmenting the pool of route candidate was investigated and although experiments showed that large route pools provided the best results, initializing the route pool

with a small set of good candidates and heuristically augmenting this set provided good solutions while enabling the solution of larger instances. It was also noted that the concept of a route pool allow planners the option of supplying the initial set of route candidates which combined with the flexibility of the heuristic route pool augmentation can provide a very high level of control over the final service network design if so desired.

The second problem that is treated in detail in this thesis is the network transition problem which is concerned with the migration of a fleet or part of a fleet from operating one service network and to operating a new and changed service network. This problem provides the link between designing a new service network and implementing/operationalizing this design and is thus fundamental for the adoption of methods for system wide redesign of the service networks such as that proposed above for the container feeder service network design problem.

The network transition problem is an extension of the pickup and delivery problem and as such generally a hard problem to solve. A parallel cooperative adaptive large neighborhood search (ALNS) heuristic was proposed to solve the network transition problem. The heuristic is fundamentally based on performing a set of neighborhood moves in the form of destroy and repair (or ruin and recreate) operations embedded in a simulated annealing framework. Several neighborhoods were proposed some of which were adapted from the pickup and delivery literature while others were specially tailored to the concrete case of liner container feeding. The ALNS framework is itself quite general and is well suited to solve highly constrained problems in which moving between feasible solutions using traditional local search based on small modifications can be difficult.

The qualities of the ALNS were analyzed through a series of computational experiments including a scalability study showing that problems of up to 400 commodities (800 requests when measured using the definition from the pickup and delivery literature) could be solved in a time frame appropriate to the application. Additional computational experiments for an adapted real-world scenario showed that the ALNS algorithm was capable of producing solutions with savings potentials of more than \$100,000 compared to the schedule currently operated by the problem owner.

Finally, adaptations of the ALNS to solve the closely related network recovery problem have been discussed thus showing a possible viable path toward obtaining a fully qualified planning system in which three central processes are captured; 1) Overall design of the service network

using the proposed liner SND model, 2) Migrating to the new service network by means of the ALNS developed for the network transition problem and finally, 3) Recovering from disruptions using an adapted version of the ALNS. Integrated, these three components could comprise a valuable planning system for liner carriers.

It is reasonable to expect that as the market continues to demand more integrated supply chains, the planning problems faced by liner carriers will only become more complex. Where operating cost and revenue were once the primary measures of the quality of a service network, carriers must now include multi-dimensional and complex objectives relating to service levels and integration to remain competitive and secure future business. Adding dimensions that are difficult to quantify and for which interactions and correlations are not well described, will lead to additional requirements to the models and tools used to address the various planning problems. Meeting such requirements means that there is a need for close collaboration between the liner carriers and the research community. This work has aimed at providing a unified description of the liner service network design problem, highlighting some of the current research challenges and possible future directions thus hopefully fueling new research in this area.

Bibliography

- R. Agarwal and E. Özlem. Ship scheduling and network design for cargo routing in liner shipping. *Transportation Science*, 42(2):175–196, 2008.
- Y. Ageeva. Approaches to incorporating robustness into airline scheduling. Master’s thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, August 2000. Report no.: ICAT-2000-8.
- R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network Flows – Theory, Algorithms and Applications*. Prentice Hall, Upper Saddle River, New Jersey, 1993.
- R. K. Ahuja, T. L. Magnanti, J. B. Orlin, and M. Reddy. Applications of network optimization. In M. O. Ball, T. L. Magnanti, C. L. Monma, and G. L. Nemhauser, editors, *Network Models*, volume 7 of *Handbook in Operations Research and Management Science*, chapter 1. Elsevier, 1995.
- Y. Almogly and O. Levin. Parametric analysis of a multi/stage stochastic shipping problem. In *Proc 5th International Conference on Operational Research (OR 69)*, pages 359–370, 1970.
- J. Andersen, T. G. Crainic, and M. Christiansen. Service network design with asset management: Formulations and comparative analyses. *Transportation Research Part C*, 17:197–207, 2009.
- A. P. Armacost. *Composite Variable Formulations for Express Shipment Service Network Design*. PhD thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 2000.

- A. P. Armacost, C. Barnhart, and K. A. Ware. Composite variable formulations for express shipment service network design. *Transportation Science*, 36(1):1–20, 2002.
- A. Balakrishnan, T. L. Magnanti, and P. Mirchandani. Network design. In M. Dell’Amico, F. Maffioli, and S. Martello, editors, *Annotated Bibliographies in Combinatorial Optimization*, Discrete Mathematics and Optimization, chapter 18, pages 311–334. John Wiley & Sons, 1997.
- C. Barnhart and R. R. Schneur. Air network design for express shipment service. *Operations Research*, 44(6):852–863, 1996.
- C. Barnhart, E. L. Johnson, G. L. Nemhauser, M. W. P. Savelsbergh, and P. H. Vance. Branch-and-price: Column generation for solving huge integer programs. *Operations Research*, 46(3):316–329, 1998.
- C. Barnhart, C. A. Hane, and P. H. Vance. Using branch-and-price-and-cut to solve origin-destination integer multicommodity flow problems. *Operations Research*, 48(2):318–326, 2000.
- C. Barnhart, P. Belobaba, and A. R. Odoni. Applications of operations research in the air transport industry. *Transportation Science*, 37(4):368–391, 2003.
- H. B. Bendall and A. F. Stent. Longhaul feeder services in an era of changing technology: an asia-pacific perspective. *Maritime Policy & Management*, 26(2):145–159, 1999.
- H. B. Bendall and A. F. Stent. A scheduling model for a high speed containership service: A hub and spoke short-sea application. *International Journal of Maritime Economics*, 3:262–277, 2001.
- G. Berbeglia, J.-F. Cordeau, I. Gribkovskaia, and G. Laporte. Static pickup and delivery problems: A classification scheme and survey. *TOP*, 15(1):1–31, 2007.
- J. R. Birge and F. c. Louveaux. *Introduction to Stochastic Programming*. Springer Series in Operations Research. Springer, 1997.
- T. B. Boffey, E. D. Edmond, A. I. Hinxman, and C. J. Pursglove. Two approaches to scheduling container ships with an application to the north atlantic route. *Journal of the Operational Research Society*, 30(5):413–425, 1979.
- N. Boland, J. Dethridge, and I. Dumitrescu. Accelerated label setting algorithms for the elementary resource constrained shortest path problem. *Operations Research Letters*, 34:58–68, 2006.

- F. Broeze. *The Globalisation of the Oceans: Containerisation from the 1950s to the Present*. Number 23 in Research in Maritime History. International Maritime Economic History Association, 2002.
- G. G. Brown, K. J. Cormican, S. Lawphongpanich, and D. B. Widdis. Optimizing submarine berthing with a persistence incentive. *Naval Research Logistics*, 44(4):301–318, 1997a.
- G. G. Brown, R. F. Dell, and R. K. Wood. Optimization and persistence. *Interfaces*, 27:15–37, 1997b.
- S.-C. Cho and A. N. Perakis. Optimal liner fleet routeing strategies. *Maritime Policy Management*, 23(3):249–259, 1996.
- M. Christiansen and K. Fagerholt. Robust ship scheduling with multiple time windows. *Naval Research Logistics*, 49:611–625, 2002.
- M. Christiansen, K. Fagerholt, and D. Ronen. Ship routing and scheduling: Status and perspectives. *Transportation Science*, 38(1):1–18, 2004. ISSN 00411655.
- M. Christiansen, K. Fagerholt, B. Nygreen, and D. Ronen. Maritime transportation. In C. Barnhart and G. Laporte, editors, *Transportation*, volume 14 of *Handbooks in Operations Research and Management Science*, chapter 4. North-Holland Publishing Company, 2007.
- E. M. Claessens. Optimization procedures in maritime fleet management. *Maritime Policy & Management*, 14:27–48, 1987.
- J. Clausen, A. Larsen, J. Larsen, and N. J. Rezanova. Disruption management in the airline industry – concepts, models and methods. *Computers & Operations Research*, 37:809–821, 2010.
- A. M. Cohn and C. Barnhart. Composite-variable modeling for service parts logistics. *Annals of Operations Research*, 144(1):17–32, 2006.
- C. E. Cortés, M. Matamala, and C. Contardo. The pickup and delivery problem with transfers: Formulation and a branch-and-cut solution method. *European Journal of Operational Research*, 200:711–724, 2010.
- T. G. Crainic and M. Gendreau. Cooperative parallel tabu search for capacitated network design. *Journal of Heuristics*, 8(6):601–627, 2002. ISSN 13811231.
- T. G. Crainic and K. H. Kim. Intermodal transportation. In C. Barnhart and G. Laporte, editors, *Transportation*, volume 14 of *Handbooks in*

- Operations Research and Management Science*, chapter chap. 8. North-Holland Publishing Company, 2007.
- T. G. Crainic, A. Fangioni, and B. Gendron. Bundle-based relaxation methods for multicommodity capacitated fixed charge network design. *Discrete Applied Mathematics*, 112:73–99, 2001.
- I. M. Datz. Planning tools for ocean transportation – ship scheduling. *Norwegian Shipping News*, 22:1064–1069, 1968.
- J. E. Davies. Competition, contestability and the liner shipping industry. *Journal of Transport Economics and Policy*, 20(3):299–312, 1986.
- EC. Optimar – benchmarking strategic options for european shipping and for the european maritime transport system in the horizon 2008-2018. Technical report, European Commission - Transport, 2008. http://ec.europa.eu/transport/strategies/2018_maritime_transport_strategy_en.htm.
- K. Fagerholt. Optimal fleet design in a ship routing problem. *International Transactions in Operational Research*, 6:453–464, 1999.
- K. Fagerholt. Ship scheduling with soft time windows: An optimisation based approach. *European Journal of Operational Research*, 131:559–571, 2001.
- K. Fagerholt. Designing optimal routes in a liner shipping problem. *Maritime Policy and Management*, 31(4):259–268, 2004.
- K. Fagerholt and H. Lindstad. Optimal policies for maintaining a supply service in the norwegian sea. *Omega*, 28(3):269–275, 2000. ISSN 03050483.
- M. L. Fisher. The lagrangian relaxation method for solving integer programming problems. *Management Science*, 27(1):1–18, 1981.
- A. Frangioni. About lagrangian methods in integer optimization. *Annals of Operations Research*, 139(1):163–193, 2005.
- M. R. Garey and D. S. Johnson. *Computers and intractability: a guide to the theory of NP-completeness*. W. H. Freeman, 1979.
- M. Gendreau, M. Iori, G. Laporte, and S. Martello. A tabu search algorithm for a routing and container loading problem. *Transportation Science*, 40(3):342–350, 2006.
- B. Gendron. Multicommodity network design test instances, 2001. URL <http://www.di.unipi.it/di/groups/optimize/Data/MMCF.html#NetDesMMCF>. <http://www.di.unipi.it/di/groups/optimize/Data/MMCF.html>.

- B. Gendron and T. G. Crainic. Relaxations for multicommodity capacitated network design problems. Technical report, Centre de recherche sur les transport, Université de Montréal, 1994. CRT-965.
- B. Gendron, T. G. Crainic, and A. Frangioni. Multicommodity capacitated network design. In B. Sansò and P. Soriano, editors, *Telecommunications Network Planning*. Kluwer Academic Publishers, 1999. Université de Montréal.
- B. Gendron, J.-Y. Potvin, and P. Soriano. A parallel hybrid heuristic for the multicommodity capacitated location problem with balancing requirements. *Parallel Computing*, 29(5):591–606, 2003. ISSN 01678191.
- C. A. Hane, C. Barnhart, E. L. Johnson, R. E. Marsten, G. L. Nemhauser, and G. Sigismondi. The fleet assignment problem: Solving a largescale integer program. *Mathematical Programming*, 70(2): 211–232, 1995.
- J. Hansen. Co2-besparelser: Co2-beregner nyt guldæg for unifeeder. *Børsen*, September 21:30, 2009.
- P. Hansen and N. Mladenović. Variable neighborhood search. In *Handbook of Metaheuristics*, International Series in Operations Research & Management Science, chapter 6, pages 145–184. Springer New York, 2003.
- M. Hersh and S. P. Ladany. Optimal scheduling of ocean cruises. *INFOR*, 27(1):48–57, 1989.
- N. Hingorani, D. Moore, and K. Tornqvist. Setting a new course in the container shipping industry. Technical report, IBM Business Consulting Services, IBM Global Services, Route 100, Somers, NY 10589, U.S.A., 2005.
- J. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms*, volume I+II. Springer-Verlag, 1993.
- K. Holmberg and D. Yuan. A lagrangean approach to network design problems. *International Transactions in Operational Research*, 5(6): 529–539, 1998.
- K. Holmberg and D. Yuan. A lagrangian heuristic based branch-and-bound approach for the capacitated network design problem. *Operations Research*, 48(3):461–481, 2000.
- S. Irnich and D. Villeneuve. The shortest-path problem with resource constraints and k-cycle eliminations for $k \geq 3$. *INFORMS Journal on Computing*, 18(3):391–406, 2006.

- D. I. Jaramillo and A. N. Perakis. Fleet deployment optimization for liner shipping, part 1. *Maritime Policy and Management*, 18(4):235–262, 1991.
- B. Kallehauge, J. Larsen, and O. B. G. Madsen. Lagrangian duality applied to the vehicle routing problem with time windows. *Computers & Operations Research*, 33:1464–1487, 2006.
- D. Kim, C. Barnhart, K. Ware, and G. Reinhardt. Multimodal express package delivery: A service network design application. *Transportation Science*, 33(4):391–407, 1999.
- K. H. Kjeldsen. Classification of routing and scheduling problems in liner shipping. Technical report, CORAL, Department of Business Studies, Aarhus School of Business, University of Aarhus, 2008.
- D. Klabjan, E. L. Johnson, G. L. Nemhauser, E. Gleman, and S. Ramaswamy. Airline crew scheduling with regularity. *Transportation Science*, 35(4):359–374, 2001.
- F. Kydland. *Simulation of Liner Operations*. Institute of Shipping Research, Bergen, 1969.
- D. E. Lane, T. D. Heaver, and D. Uyeno. Planning and scheduling for efficiency in liner shipping. *Maritime Policy and Management*, 14: 109–125, 1987.
- S. A. Lawrence. *International Sea Transport: The Years Ahead*. Lexington Books, 1972.
- L. H. Lee, E. P. Chew, and M. S. Sim. A heuristic to solve a sea cargo revenue management problem. *OR Spectrum*, 29:123–136, 2007.
- M. Levinson. *The Box*. Princeton University Press, 1st paperback printing edition, 2006.
- H. R. Lourenço, O. C. Martin, and T. Stützle. Iterated local search. In F. Glover and G. Kochenberger, editors, *Handbook of Metaheuristics*, chapter 11, pages 321–353. Springer, 2003.
- T. L. Magnanti and R. T. Wong. Network design and transportation planning: Models and algorithms. *Transportation Science*, 18(1):1–55, 1984.
- R. Martí. Multi-start methods. In *Handbook of Metaheuristics*, volume 57 of *International Series in Operations Research & Management Science*, chapter 12, pages 355–368. Springer New York, 2003.

- M. Minoux. Network synthesis and optimum network design problems: Models, solution methods and applications. *Networks*, 19(3):313–360, 1989.
- S. Mitrović-Minić and G. Laporte. The pickup and delivery problem with time windows and transshipment. *INFOR*, 44(3):217–227, 2006.
- S. Murray. Green supply chains: moving beyond logistics. *Financial Times*, online, Dec. 10th 2007. Dec. 10th.
- A. Nedić. *Subgradient Methods for Convex Minimization*. PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 2002.
- T. Notteboom and J.-P. Rodrigue. Containerisation, box logistics and global supply chains: The integration of ports and liner shipping networks. *Maritime Economics & Logistics*, 10:152–174, 2008.
- T. E. Notteboom and B. Vernimmen. The effect of high fuel costs on liner service configuration in container shipping. *Journal of Transport Geography*, 17:325–337, 2009.
- C. A. Olson, E. E. Sorenson, and W. J. Sullivan. Medium-range scheduling for a freighter fleet. *Operations Research*, 17(4):565–582, 1969.
- A. C. Paixão and P. B. Marlow. Strengths and weaknesses of short sea shipping. *Marine Policy*, 26:167–178, 2002.
- S. N. Parragh, K. F. Doerner, and R. F. Hartl. A survey on pickup and delivery problems - part i. *Journal für Betriebswirtschaft*, 58(1):21–51, 2008a.
- S. N. Parragh, K. F. Doerner, and R. F. Hartl. A survey on pickup and delivery problems - part ii. *Journal für Betriebswirtschaft*, 58(2): 81–117, 2008b.
- M. B. Pedersen, T. G. Crainic, and O. B. G. Madsen. Models and tabu search metaheuristics for service network design with asset-balance requirements. *Transportation Science*, 43(2):158–177, 2009.
- A. N. Perakis and D. I. Jaramillo. Fleet deployment optimization for liner shipping, part 1. *Maritime Policy and Management*, 18(3):183–200, 1991.
- B. J. Powell and A. N. Perakis. Fleet deployment optimization for liner shipping: an integer programming model. *Maritime Policy and Management*, 24(2):pp. 183–192, 1997.

- K. Rana and R. G. Vickson. A model and solution algorithm for optimal routing of a time-chartered containership. *Transportation Science*, 22(2):83–95, 1988.
- K. Rana and R. G. Vickson. Routing container ships using lagrangean relaxation and decomposition. *Transportation Science*, 25(3):pp. 201–214, 1991.
- N. J. Rezanova. *The Train Driver Recovery Problem – Solution Method and Decision Support System Framework*. PhD thesis, Technical University of Denmark, Department of Management Engineering, 2009.
- C. Rocha, B. Jaumard, and T. Stidsen. A hierarchical decomposition method for efficient computation of path-protecting p-cycles. *Telecommunication Systems*, TBD, 2009.
- D. Ronen. Cargo ships routing and scheduling: Survey of models and problems. *European Journal of Operational Research*, 12(2):pp. 119–126, 1983.
- D. Ronen. Ship scheduling: The last decade. *European Journal of Operational Research*, 71(3):pp. 325–333, 1993.
- K. Shintani, A. Imai, E. Nishimura, and S. Papadimitriou. The container shipping network design problem with empty container repositioning. *Transportation Research Part E*, 43:39–59, 2007.
- M. M. Sigurd, N. L. Ulstein, B. Nygreen, and D. M. Ryan. Ship scheduling with recurring visits and visit separation requirements. In G. Desaulniers, J. Derosiers, and M. M. Solomon, editors, *Column Generation*, GERAD 25th anniversary series, chapter 8, pages 225–245. Springer Science, 2005.
- K. R. Smilowitz, A. Atamtürk, and C. F. Daganzo. Deferred item and vehicle routing within integrated networks. *Transportation Research Part E*, 39:305–323, 2003.
- B. C. Smith and E. L. Johnson. Robust airline fleet assignment: Imposing station purity using station decomposition. *Transportation Science*, 40(4):497–516, 2006.
- H. Stadtler and C. Kilder, editors. *Supply chain management and advanced planning : Concepts, models, software and case studies*. Springer, 3rd edition, 2005.
- P. Toth and D. Vigo, editors. *The Vehicle Routing Problem*. Discrete Mathematics and Applications. SIAM, 2002.

- M. Transmodal. Far east trade cargo analysis. *Containerisation International*, 42(2):4–5, February 2009a.
- M. Transmodal. European tradelane cargo analysis. *Containerisation International*, 42(4):4–5, April 2009b.
- UNCTAD. Review of maritime transport 2008. Technical report, United Nations Conference on Trade and Development, Geneva, 2008.
- UNCTAD. Review of maritime transport 2009. Technical report, United Nations Conference on Trade and Development, Geneva, 2009.
- B. Vernimmen, W. Dullaert, and S. Engelen. Schedule unreliability in liner shipping: Origins and consequences for the hinterland supply chain. *Maritime Economics & Logistics*, 9:193–213, 2007.
- F. Weldon, editor. *Operational Simulation of a Freighter Fleet*, Research Techniques in Maritime Transportation, Washington, 1959. National Research Council publication 720.
- F. L. Weldon. Cargo containerization in the west coast-hawaiian trade. *Operations Research*, 6(5):649–670, 1958.
- R. Wright. Engines of globalisation: the story of maersk. Financial Times, October 2nd 2006.
- WTO. World trade report 2008 — trade in a globalizing world. Technical report, World Trade Organization, 2008.

A Lagrangian Relaxation Based Approach to Balanced Network Design

Section 4.1 discussed the use of a multicommodity capacitated network design (MCND) model in the context of service network design. Several limitations were noted both with respect to modeling the service network design problem but also in terms of actually solving the MCND using standard methods as linear programming (LP) and branch and bound. Despite these challenges, an extended version of the MCND was proposed in which the design should balance in all nodes, i.e., the overall network design should be decomposable as a series of simple cycles. This model was referred to as the balanced multicommodity capacitated network design (BMCND) model (refer to section 4.2 for a more elaborate definition and discussion of the relationship with the MCND).

As the BMCND extends the MCND and the new constraints do not immediately improve the strength of the formulation, it is clear that there are significant challenges associated with solving the BMCND using standard methods. Alternative approaches such as heuristics (Pedersen et al. (2009) propose a tabu search) can provide a viable alternative to standard linear programming and branch and bound. However, although typically quite fast and capable of solving large problem instances, these

methods generally do not provide any performance guarantees. This chapter instead investigates whether some of the successful applications of Lagrangian based relaxation (see e.g., Fisher (1981) for an introduction) on the MCND problem can be transferred to the balanced version of this same problem, the BMCND. For applications of Lagrangian relaxation to the MCND refer to e.g., Holmberg and Yuan (1998, 2000), Crainic et al. (2001).

A.1 Relaxation of the BMCND

For the sake of reference, the balanced multicommodity capacitated network design (BMCND) model of chapter 4 is repeated below.

Let \mathcal{A} denote a set of arcs, \mathcal{N} a set of nodes in a graph $G = (\mathcal{N}, \mathcal{A})$. Furthermore, let \mathcal{K} denote a set of commodities that must be transported over this graph, each having an associated origin $O(k)$ and destination $D(k)$, $k \in \mathcal{K}$ defining the end points of possible transport paths. Additionally, using standard notation, let $y_{ij} = 1$ if arc $(i, j) \in \mathcal{A}$ is open for service and $y_{ij} = 0$ otherwise and let x_{ij}^k denote the amount of commodity $k \in \mathcal{K}$ flowing on arc $(i, j) \in \mathcal{A}$. Finally, letting u_{ij} denote the capacity of arc (i, j) , d^k the demand of commodity $k \in \mathcal{K}$ with d_i^k denoting the demand at node $i \in \mathcal{N}$. Also define $b_{ij}^k = \min(u_{ij}, d^k)$ and the cost of operating service on an arc $(i, j) \in \mathcal{A}$ denoted by f_{ij} and similarly c_{ij}^k denotes the cost of flowing one unit of commodity k on arc (i, j) . With these definitions the balanced multicommodity capacitated network design problem can be expressed as follows:

$$(\text{BMCND}) \quad \min \sum_{(i,j) \in \mathcal{A}} f_{ij} y_{ij} + \sum_{k \in \mathcal{K}} \sum_{(i,j) \in \mathcal{A}} c_{ij}^k x_{ij}^k \quad (\text{A.1})$$

s.t.

$$\sum_{j:(i,j) \in \mathcal{A}} x_{ij}^k - \sum_{j:(j,i) \in \mathcal{A}} x_{ji}^k = d_i^k \quad \forall i \in \mathcal{N}, k \in \mathcal{K} \quad (\text{A.2})$$

$$(\alpha_{ij}) \quad \sum_{k \in \mathcal{K}} x_{ij}^k \leq u_{ij} y_{ij} \quad \forall (i,j) \in \mathcal{A} \quad (\text{A.3})$$

$$(\beta_{ij}^k) \quad x_{ij}^k \leq b_{ij}^k y_{ij} \quad \forall (i,j) \in \mathcal{A}, k \in \mathcal{K} \quad (\text{A.4})$$

$$\sum_{j:(i,j) \in \mathcal{A}} y_{ij} - \sum_{j:(j,i) \in \mathcal{A}} y_{ji} = 0 \quad \forall i \in \mathcal{N} \quad (\text{A.5})$$

$$y_{ij} \in \{0, 1\} \quad \forall (i,j) \in \mathcal{A} \quad (\text{A.6})$$

$$x_{ij}^k \in \mathbb{R}_+ \quad \forall (i,j) \in \mathcal{A}, k \in \mathcal{K} . \quad (\text{A.7})$$

The objective is to minimize the total fixed and variable costs of satisfying all commodities while respecting capacity (constraints (A.3) and (A.4)) and ensuring design balance (constraints (A.5)).

It is observed that constraints (A.3) and (A.4) are the only two constraints linking the design (y_{ij}) and flow (x_{ij}^k) variables. Introducing the non-negative Lagrange multipliers $\alpha_{ij} \geq 0$ and $\beta_{ij}^k \geq 0$, $(i,j) \in \mathcal{A}, k \in \mathcal{K}$ for constraints (A.3) and (A.4) respectively and relaxing these in a Lagrangian fashion we obtain the Lagrangian relaxed (LR) problem below.

$$(LR1) \tag{A.8}$$

$$z_D(\alpha, \beta) = \min \sum_{(i,j) \in \mathcal{A}} \left(f_{ij} - \alpha_{ij} u_{ij} - \sum_{k \in \mathcal{K}} \beta_{ij}^k b_{ij}^k \right) y_{ij} \\ + \sum_{k \in \mathcal{K}} \sum_{(i,j) \in \mathcal{A}} (c_{ij}^k + \alpha_{ij} + \beta_{ij}^k) x_{ij}^k \tag{A.9}$$

s.t.

$$\sum_{j: (i,j) \in \mathcal{A}} x_{ij}^k - \sum_{j: (j,i) \in \mathcal{A}} x_{ji}^k = d_i^k \quad \forall i \in \mathcal{N}, k \in \mathcal{K} \tag{A.10}$$

$$\sum_{j: (i,j) \in \mathcal{A}} y_{ij} - \sum_{j: (j,i) \in \mathcal{A}} y_{ji} = 0 \quad \forall i \in \mathcal{N} \tag{A.11}$$

$$y_{ij} \in \{0, 1\} \quad \forall (i, j) \in \mathcal{A} \tag{A.12}$$

$$x_{ij}^k \in \mathbb{R}_+ \quad \forall (i, j) \in \mathcal{A}, k \in \mathcal{K} . \tag{A.13}$$

As LR1 is a relaxation of BMCND it is clear that $z_D(\alpha, \beta) \leq z$ and thus, $z_D(\alpha, \beta)$ provides a lower bound on the objective of BMCND for any values of α and β representing the vectorized form of α_{ij} and β_{ij}^k , $(i, j) \in \mathcal{A}, k \in \mathcal{K}$ respectively. In fact, since LR1 possesses the integer property (Fisher, 1981), the lower bound provided by LR1 is equivalent to the liner programming relaxation of BMCND.

It is immediatly seen that (LR1) decomposes into \mathcal{K} shortest path problems (constraints (A.10) and (A.13) and the last term of (A.9)) and a minimum cost circulation problem (constraints (A.11)–(A.12) and the first term of (A.9)). We observe that the coefficient matrix of the circulation problem is totally unimodular which combined with integer right hand sides means that an integer solution is obtained from the LP relaxation. Thus, both sub-problems are easily solved and posses the integrality property. Although this fact means that the bounds obtained from (LR1) are equivalent to those obtained by the LP relaxation of BMCND (which is known to be quite weak) we may exploit the primal information (as suggested by (Frangioni, 2005)) obtained from LR1 to build a heuristic which will serve both as a way to obtain upper bounds for pruning in the branch and bound tree as well as intermediate primal solutions. Furthermore, for large instances, it may be more efficient to solve LR1 to obtain a lower bound than solving the LP relaxation of BMCND problem

containing many more constraints.

The Lagrangian problem LR1 may be augmented by observing that the total flow on any arc can never exceed the capacity on the arc thus leading to the additional constraint set

$$\sum_{k \in \mathcal{K}} x_{ij}^k \leq u_{ij} \quad \forall (i, j) \in \mathcal{A} \quad (\text{A.14})$$

In the original formulation (BMCND) this constraint set will be dominated by the weak forcing constraints (A.3). However, in the relaxed problem (LR1) these constraints may be added thus integrating the $|\mathcal{K}|$ shortest path problems into a single capacitated multicommodity minimum cost flow (MMCF) problem. Although this problem is more complicated to solve than a series of shortest path problems, it can lead to the faster discovery of feasible solutions in a branch and bound search. The problem consisting of equations (A.9) – (A.13) with the addition of equation (A.14) will be referred to as (LR2).

It is possible to alternate between solving LR1 and LR2 at certain points of the B&B search. In particular, as more arcs are fixed to zero in the B&B tree, the multicommodity flow problem in LR2 becomes smaller as certain variables can be eliminated.

A.2 The Lagrangian Dual Problem

Obviously, when used in a branch and bound setup, the goal is to determine the values of α and β such that $z_D(\alpha, \beta)$ is maximal. The resulting problem is referred to as the Lagrangian dual problem and can be expressed more formally as

$$z_{LD} = \max_{(\alpha, \beta)} Z_D(\alpha, \beta) . \quad (\text{A.15})$$

Traditionally, the solution of the Lagrangian dual A.15 has been approached by means of the classical subgradient method. In this method, a subgradient $s_i(\alpha, \beta)$ of the Lagrangian relaxed problem evaluated at a point (α, β) is used to determine the direction in which to move and a step size t_i determines how far to move in that direction. For a fixed solution (x', y') to the Lagrangian relaxed problem evaluated at (α, β) the subgradient vector s_i at iteration i is given as $s_i = (s_i^1, s_i^2, \dots, s_i^n) \in \mathbb{R}^n$, $n = (1 + |\mathcal{K}|) \cdot |\mathcal{A}|$ with the first $|\mathcal{A}|$ elements given by

$$s_i^a = \sum_{k \in \mathcal{K}} x_{ij}^k - u_{ij} y_{ij}, \quad a = (i, j) \in \mathcal{A}, \quad (\text{A.16})$$

and the following $|\mathcal{K}| \cdot |\mathcal{A}|$ elements given by

$$s_i^{ak} = x_{ij}^k - b_{ij}^k y_{ij}, \quad a = (i, j) \in \mathcal{A}, k \in \mathcal{K}. \quad (\text{A.17})$$

The subgradient method is initialized with an initial value of $\lambda_1 = (\alpha_1, \beta_1)$ and at each iteration, the next Lagrangian multipliers are given as

$$\lambda_{i+1} = \lambda_i + t_i \cdot s_i. \quad (\text{A.18})$$

The process is continued until an iteration limit or a tolerance is met. The major appeal of this method is the implementational simplicity and low complexity in evaluation. However, setting the step size t_i of the subgradient method is very difficult and convergence is generally unstable and highly dependent on this setting (see Nedić (2002) for an analysis of several subgradient methods).

An alternative to the simple subgradient method is the use of cutting plane method to represent a model of the Lagrangian dual problem (A.15). Rather than simply using a single subgradient in each iteration, the idea behind the cutting plane method is to accumulate the generated subgradients to get a model of the Lagrangian dual problem expressed as follows

$$\begin{aligned} z_{LD} &= \max w \\ \text{s.t.} \end{aligned} \quad (\text{A.19})$$

$$w \leq z_D(\lambda_i) + s_i^T (\lambda - \lambda_i) \quad i = 1, \dots, q \quad (\text{A.20})$$

$$(\text{A.21})$$

Unfortunately, cutting plane method is known to converge quite slowly and be quite unstable as it tends to result in extreme values of λ when the number of cutting planes q provide a poor model of z_{LD} . To overcome this problem, the cutting plane method can be stabilized through the introduction of a trust-region (Hiriart-Urruty and Lemaréchal, 1993, vol. II, chap. XV) constraining the allowed range of the Lagrangian multipliers λ . Defining the trust region as a simple box around a stability center μ_q the cutting plane model (A.20) can be extended with the following additional stabilizing constraints

$$\mu_q - \Delta_q \leq \lambda \leq \mu_q + \Delta_q, \quad (\text{A.22})$$

The size of the trust region Δ_q at iteration q is updated based on the progress in the objective of the Lagrangian relaxed problem $z_D(\lambda_q)$ relative to that predicted by the model (A.20) defined as follows

$$\rho = \frac{z_D(\lambda_{q+1}) - z_{LD}(\mu_q)}{z_{LD}(\lambda_{q+1}) - z_{LD}(\mu_q)}. \quad (\text{A.23})$$

If ρ is equal to 1, the cut obtained in the previous iteration did not contribute with any new information and thus the cutting plane model of the Lagrangian dual problem describes z_{LD} accurately within the trust region. Thus, the size of the trust region is increased with a factor of $\tau^+ > 1$ to allow the cutting plane model to discover new constraint more quickly. On the other hand, if $\rho < 0$, a step has been taken into a region where the cutting plane model is a poor representation of z_{LD} and thus, the trust region size is decreased with a factor of $0 < \tau^- < 1$ to discover new constraints and avoid algorithm stalling. This update mechanism was originally proposed by Kallehauge et al. (2006). At each iteration, the stability center of the trust region is moved if $\rho \geq \underline{\text{delta}}_\rho$, where $\underline{\text{delta}}_\rho$ is the improvement tolerance for a move.

We express the overall flow of the cutting plan algorithm as follows:

Step 0 (Initialization): Choose initial values for the multipliers $\mu_1 = \lambda_1 = (\alpha_1, \beta_1)$, initial trust region width Δ_1 , iteration counter $i = 1$, and stopping tolerance $\underline{\delta}$. Solve $z_D(\lambda_s)$ and compute the subgradient $s_1 = s(\lambda_1)$.

Step 1 (Lagrange dual problem): Solve the restricted Lagrangian dual problem

$$z_{LD}(\lambda) = \max z \quad (\text{A.24})$$

$$\text{subject to} \quad (\text{A.25})$$

$$z \leq z_D(\lambda_i) + s_i^T(\lambda - \lambda_i) \quad i = 1, \dots, q \quad (\text{A.26})$$

$$\mu_q - \Delta_q \leq \lambda \leq \mu_q + \Delta_q \quad (\text{A.27})$$

to get the solution λ_{q+1} and calculate $\delta_q = z_{LD}(\lambda_{q+1}) - z_D(\mu_q)$.

Step 2 (Stopping criterion): Stop if $\delta_q \leq \underline{\delta}$.

Step 3 (Lagrangian relaxed): Solve $z_D(\lambda_{q+1})$ and compute the new subgradient $s_{q+1} = s(\lambda_{q+1})$.

Step 4 (Update): Determine ρ . If $\rho = 1$ update $\Delta_{q+1} = \Delta_q * \tau^+$, else if $\rho \leq 0$ update $\Delta_{q+1} = \Delta_q * \tau^-$, else if $0 < \rho < 1$ update $\Delta_{q+1} = \Delta_q$.

Finally, if $\rho \geq \delta_\rho$ update $\mu_{q+1} = \lambda_{q+1}$.
Continue at step 1.

Initially, $\lambda_0 = \mathbf{1}$, a vector of all ones, is used as the starting value for the Lagrangian multipliers.

A.3 Heuristic and Branch-and-Bound

Some very preliminary work has been done embedding the Lagrangian based lower bounding procedure in a parallel branch and bound framework developed specifically for this problem. The branch selection strategy is best-first and branching is performed on the original design variables such that constraints $y_{ij} = 0$ and $y_{ij} = 1$ are imposed in a new nodes. At each new branching step, the variable for the most constraining arc $(i, j) \in \mathcal{A}$ is chosen as the branching variable based on the Lagrange multipliers. Formally, the next branching variable y_{ij} is the one for which arc (i, j) satisfies

$$(i, j) = \arg \max_{(i, j) \in \mathcal{A}} (\alpha_{ij} + \sum_{k \in \mathcal{K}} \beta_{ij}^k) . \quad (\text{A.28})$$

Each branching decision imposes additional constraints on the sub-problems. In the circulation problem, constraints can be handled through variable fixing. For the shortest path problems, constraints can be captured through appropriately setting costs on closed arcs (the $y_{ij} = 0$ branch) and verifying feasibility. In the capacitated multicommodity minimum cost flow problem, branches are handled through appropriate definition of the right hand sides of (A.14).

The framework is still at a relatively early stage and more work is needed to verify the implementation as well as the integration of a primal heuristic to provide upper bounds during the search and help speed up the branch and bound search.

A.4 Preliminary Computational Experiments

As this is still work in progress, results are so far limited and no clear conclusions can be made at this point. Experiments on small instances with only ten commodities show that the above proposed solution method takes around twice as long to solve the root node when compared to solving the LP relaxation of BMCND using CPLEX 11.1. Similarly,

solving the problem to optimality using the simple branch and bound algorithm proposed in the previous section without any primal heuristic takes around 40 times as long as using CPLEX 11.1 with standard settings. These results are not encouraging although a primal heuristic is believed to greatly improve the performance of the branch and bound procedure.

Looking at larger instances with up to 200 commodities, the results change slightly. In these cases, the Lagrangian relaxation based algorithm initially converges faster in the root node than the dual simplex algorithm of CPLEX. However, problems remain with ensuring progress of the cutting plane method as it converges towards the optimal value of the Lagrangian relaxed problem ultimately becoming outperformed by the dual simplex algorithm. At this point, experiments suggest that some improvements can be realized through appropriate adjustment of algorithm parameters although it is not clear whether this will lead to an overall performance of the proposed cutting plane method that is competitive with state of the art dual simplex implementations.

In a more broad perspective, it remains interesting to investigate how well the above proposed solution method performs when augmented with a primal heuristic. Combined with a branch and bound search and early termination of the cutting plane algorithm, this could lead to an algorithm capable of producing good solutions in a time frame competitive to standard mixed integer program solvers.

APPENDIX B

Data Generator

Working with real-world data in problems based on industry applications can be complicated by several factors. The primary challenge is simply *availability* as procuring real-world data of a sufficient quality and with realistic properties can be very difficult in academic projects. Typically, the availability of data depends on the presence of an industry partner and requirements to data quality and attributes may mean that a significant amount of work is put on this partner to extract and filter data. Even if data are readily available, research typically requires some control over data to achieve certain properties necessary to evaluate a certain algorithm under a diverse set of conditions. In order to obtain sufficient control over data

Although there are many advantages of working with synthetic data, their use does pose one significant challenge which is hard to overcome. Typically, the performance of real-world networks is evaluated based on a predefined set of so called key performance indicators. In liner shipping, this could be revenue per voyage or average capacity utilization. Creating synthetic data that accurately captures detailed properties of a service network that again translate into satisfactory levels for the performance indicators is very difficult. This means that while synthetic data can be used to evaluate algorithmic performance and behaviour, absolute evaluations of performance indicators should be avoided. Instead, analysis based on synthetic data should focus on relative evaluations comparing

algorithmic performance subject to varying one or more overall properties of a problem instance.

Instance properties can be adjusted and controlled through an extensive set of input parameters determining the behaviour of the data generator. Parameters fall in two categories; fixed and interval, where the actual value of the property controlled by an interval variable is selected at random from a uniform distribution in the specified interval. It is possible to extend the implementation to allow the generator to bias ranged attributes towards a certain end of the interval. The relevant parameters of the data generator have been summarized in section B.4 for the sake of reference. The following sections assume the parameter set to be given and consistent (in terms of both units and scale).

While the data generator has been developed for the network transition problem discussed in chapter 6, it can be adapted to generate data for a generalized form of a pickup and delivery problem with a fixed vehicle fleet.

The following sections will elaborate on the principles used in the generation of instances for the network transition problem. The presentation is divided into three sections describing the geography, the commodities, and the vessel schedules respectively.

B.1 Geography

Ports define the geographical layout of a service network and represent the nodes at which freight (containers) are handled in this network. Main geography is constrained to a closed two-dimensional Euclidean plane with boundaries specified by $[0; X_{\text{Max}}] \times [0; Y_{\text{Max}}]$ and may contain a sub-region defined by $[0; X_{\text{MaxHubRegion}}] \times [0; Y_{\text{MaxHubRegion}}]$ with $X_{\text{MaxHubRegion}} < X_{\text{Max}}$ and $Y_{\text{MaxHubRegion}} < Y_{\text{Max}}$. Port locations (given as (x_A, y_A) for a port A) are determined from two discrete uniform distributions with interval ends defined by the boundaries of the geography. Overall port distribution falls in two categories. For the *random* instances identified by $\text{NumHubs} = 0$, NumPorts ports are distributed over the entire main geography ignoring any sub-region (setting $X_{\text{MaxHubRegion}} = Y_{\text{MaxHubRegion}} = 0$). For the *cluster* instances ($\text{NumHubs} > 0$), a number (NumHubs) of hub ports are located in the sub-region and furthermore, $\text{NumPorts} - \text{NumHubs}$ non-hub ports (also called outports) are located in the main geography, excluding the sub-region. The cluster instances are designed to mimic a concrete case in a short-sea shipping application where hubs are located in a cluster in one

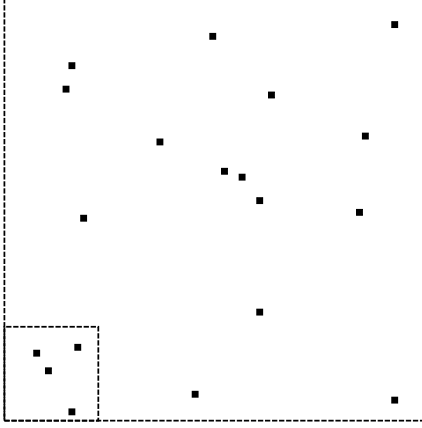


Figure B.1: Example of geographical port layout for a problem with a hub region (bottom left).

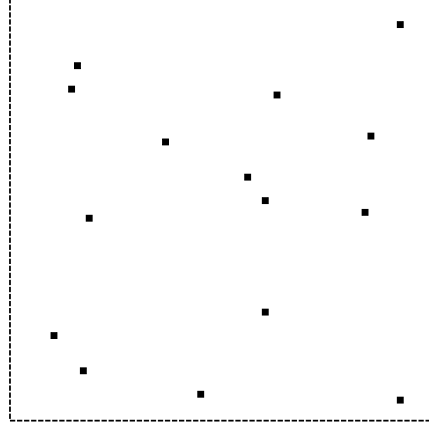


Figure B.2: Example of geographical port layout for a problem that does not contain a hub region.

extreme of the network of ports.

Examples of a random and a cluster instance are illustrated in Figure B.1 and Figure B.2 respectively.

Associated with each port is a transshipment cost per TEU as well as a berthing cost that is paid each time a vessel visits the port. There are no time windows specifying when the port is open for service and it is assumed that transshipment and berthing costs are independent of the time at which a port is visited.

A distance table can be derived from the port locations by simply calculating the Euclidean distance between any two ports (with the distance between two ports A and B defined as $d_{AB} = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$). The absence of an explicit distance table means that the data generator does not currently cater for the inclusion of alternative routes between ports such as those offered by canals.

B.2 Commodities

Commodities are generally the driving force of many routing and scheduling problems. However, generating random data instances with artificial demand patterns attempting to mimic real-world behaviour is very dif-

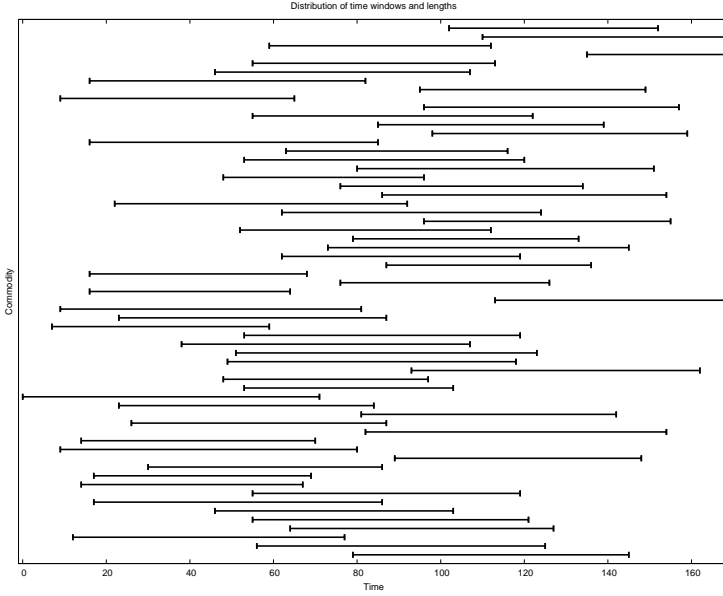


Figure B.3: Distribution of time windows for a test instance with a period of 168 hours. Average travel time between ports is ~ 23 hours.

ficult. The data generator implements some degree of control over the demand pattern as will be described below, but in general, the generated patterns will be more complicated and random than the real-world patterns they seek to emulate.

A commodity is defined by distinct pickup and a delivery locations and a volume (measured in TEU in the current application). For each commodity, a time window is generated specifying when it is available for pickup but no time window is generated for the delivery. Algorithmically, the delivery time window can be derived from the pickup TW combined with the distance and problem time horizon information. Figure B.3 shows an example of how time windows are distributed for random instance.

With each commodity, there is an associated unit revenue drawn at random from a predefined interval $[\text{RevenueTeuMin}; \text{RevenueTeuMax}]$. While it may be argued that the revenue should depend on the distance between the pickup and delivery location of the commodity, this is far from always the case in real-world applications. Multiple aspects not modelled in the data generator will impact the price (and thus the required revenue) of transporting a single unit between a pickup and a delivery location, in-

cluding the trade balance at these two locations and vessel utilization. An example of this price imbalance is Far East – Europe trade lane where vessels are mostly full going out of the Far East, but typically carry less freight on the return trip leading to a price for Far East – Europe being several times higher than in the west-bound direction.

Generally, the number of commodities is much larger than the number of physical locations at which these are to be picked up/delivered. To control the total number of commodities generated, an upper limit can be imposed through the parameter `NumCommoditiesMax`. However, in order to facilitate the definition of instance properties relating to capacity constraints, a measure of teu-hours has been introduced as a way to express the total system capacity. The total teu-hours capacity is calculated as the sum over all vessels of vessel capacity times the number of hours the vessel is available between start and end of the transition period. Each commodity consumes an amount of teu hours proportional to its volume (TEU) and an estimate of the time it is expected to spend on a vessel from pickup to delivery. An additional parameter `ComJourneyLength` controls the time estimate and allows for the creation of instances that are tightly or loosely capacity constrained. Commodities are generated until the maximum limit is reached or the total amount of available teu-hours has been consumed. Although this approach does not explicitly take geographical complexities into account, it greatly simplifies the process of creating valid instances.

For the random instances, pickup and delivery locations are selected randomly for all the generated commodities. In cluster instances, commodity flow is classified as either hub→non-hub (or non-hub→hub) or non-hub→non-hub depending on the type of the pickup and delivery ports of individual commodities. Overall commodity flow can be controlled such that the distribution (measured as the total amount of teu-hours) between the two categories satisfies a fixed value (`ComDistribHubNonhubPct`). Again, this is to mimic behaviour in a concrete short-sea shipping application where as much as 90% of the flow is hub→non-hub (or non-hub→hub).

B.3 Schedules

Schedule generation falls in two different categories; fixed and transition schedules. Fixed schedules are generated for the whole planning horizon including the transition period. The schedule for non-transitioning (fixed) vessels is generated by first selecting an ordered set of ports to

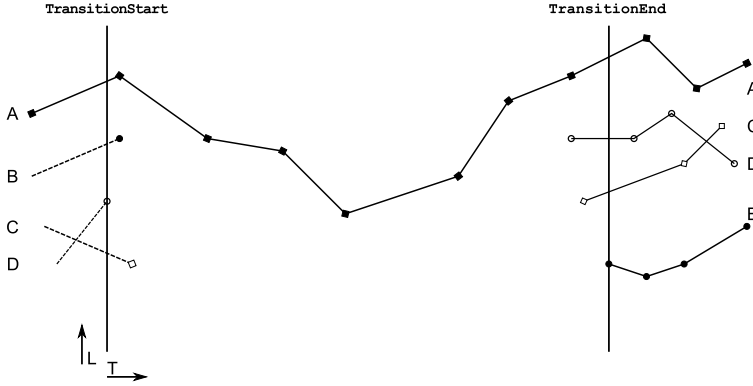


Figure B.4: Example of schedules generated for one fixed vessel (A) and three transitioning vessels (B, C, and D). Location varies along the vertical axis and time along the horizontal axis.

be visited by that vessel. The total number of ports in this set determined by the parameters `PortCallsMin` and `PortCallsMax` and the set order represents the sequence in which ports are visited. Next, a start time (and date) for this fixed schedule is selected in the interval from `ProblemStart` to `TransitionStart`. A cyclic schedule is created by first attaching the schedule start time to the first port in the port set and then continuing assigning port call times for the remaining ports in the port set using distance and port stay time (selected in `[PortCallTimeMinHours; PortCallTimeMaxHours]`) to determine these times. This process continues to roll the schedule forward in time repeating the port sequence until a port call time exceeds the `ProblemTimeEnd` time limit. Example: Port sequence is A, B, G, I, C, start time is 0 and distances are $(A, B) = 2$, $(B, G) = 3$, $(G, I) = 1$, $(I, C) = 4$, $(C, A) = 4$. With `ProblemTimeEnd` = 20 and ignoring port stay time, the schedule becomes A(0), B(2), G(5), I(6), C(10), A(14), B(16), and G(17) with A(0) interpreted as visiting A at time 0.

Transition schedules consist of port calls either belonging to the old or the new schedule. Generation of transition schedules proceeds as for the fixed schedules with the exception that no port calls are scheduled within the transition period except for a single port call from the old schedule at the beginning of the transition period and the first port call of the new schedule at the end of the transition period. The two port calls within the transition period are the fix-points of the vessel schedule that is to be determined and can be thought of as depots in traditional vehicle routing terms. Figure B.4 illustrates an example of

how a complete instance schedule can look when there is a single vessel on a fixed schedule and three transitioning vessels.

B.4 Parameter Overview

As mentioned above, the behaviour of the data generator is controlled through a set of parameters defining the properties of the generated test cases. Below is a brief summary of these parameters:

Parameter	Purpose
	<i>Size and Period</i>
ProblemStart	Earliest date from which schedule information is generated
ProblemEnd	Latest date to which schedule information is generated
TransitionStart	Cut-off date for the start of the transition period
TransitionEnd	Cut-off date for the end of the transition period
NumCommoditiesMax	Maximum number of commodities to generate
NumFixedVessels	Number of non-transitioning vessels (schedules)
NumHubs	Number of ports designated as hubs
NumPorts	Total number of ports (including hubs)
NumTransitioningVessels	Number of transitioning vessels
XMax	Maximum geographical size of the problem region in the x-axis direction (units should agree with VesselSpeed)
YMax	As above for the y-axis
XMaxHubRegion	As above for the hub region for the x-axis
YMaxHubRegion	As above for the y-axis

Parameter	Purpose
<i>Ranged Attributes</i>	
BunkerCostNmiMin	Lower limit for the cost of fuel (bunker) per nautical mile
BunkerCostNmiMax	Upper limit for the cost of fuel
ComJourneyLength	Fraction of the time from availability to the end of the planning horizon a commodity consumes capacity (TEU-Hours)
ComDistribHubNonhubPct	Fraction of demand traveling hub→non-hub or non-hub→hub
CommodityDelayMin	Minimum time window length for the availability of a commodity
CommodityDelayMax	Maximum time window length (see previous)
CommodityTeuMin	Minimum absolute commodity volume in TEU
CommodityTeuMaxPct	Maximum commodity volume (TEU) relative to the capacity of the smallest vessel
DailyTcCostMin	Minimum daily vessel time charter cost
DailyTcCostMax	Maximum daily vessel time charter cost
PortCallTimeMinHours	Minimum limit for the port time included in the generated port calls (schedule)
PortCallTimeMaxHours	Maximum limit for the port time (see previous)
RevenueTeuMin	Minimum revenue per TEU
RevenueTeuMax	Maximum revenue per TEU
TransshipmentCostMin	Minimum transshipment cost per TEU
TransshipmentCostMax	Maximum transshipment cost per TEU
VesselSpeedMin	Minimum vessel speed (in nautical miles)
VesselSpeedMax	Maximum vessel speed
VesselTeuMin	Minimum vessel TEU capacity
VesselTeuMax	Maximum vessel TEU capacity
<i>Miscellaneous</i>	
RandomSeed	Random seed input allowing reproducible results for a fixed set of parameters. -1 seeds the random number generator with the current time.

Glossary and Abbreviations

BMCND	Balanced Multicommodity Capacitated Network Design
Bunker	Equivalent to fuel in maritime applications
Consignee	The receiver of shipped/transported goods
Fairway fee	A fee specific to a certain vessel visiting a given port. The fee may e.g., be payed for the first 10 visits after which no more fees are payed.
FEU	Fourty-foot Equivalent Unit (2 TEU)
Knot	Speed measure. 1 knot = 1 nautical mile per hour
MCND	Multicommodity Capacitated Network Design
MLBMCND	Multi-Layer Balanced Multicommodity Capacitated Network Design
NVOCC	Non-Vessel Operating Common Carrier
Reefer	A refrigerated ship designed to carry (palletized) perishable goods. Also used as short for a refrigerated container.

RO-RO	Roll-on/roll-off vessel for the transport of wheeled cargo units.
Rotation	A closed cycle of port visits performed by one or more vessels
Shipper	The sender of goods
Slow steaming	Sailing at a speed slower than the nominal.
SND	Service Network Design
TEU	Twenty-foot Equivalent Unit
Voyage	The journey between two ports. Occasionally used to distinguish multiple iterations of a rotation.

DTU Transport performs research and provides education on traffic and transport planning. It advises the Danish Ministry of Transport on infrastructure, economic appraisals, transport policy and road safety and collects data on the transport habits of the population. DTU Transport collaborates with companies on such topics as logistics, public transport and intelligent transport systems.

DTU Transport
Department of Transport
Technical University of Denmark

Bygningstorvet 115
2800 Kgs. Lyngby
Tel. +45 45 25 15 24
Fax +45 45 93 64 12

www.transport.dtu.dk