Université de Montréal

Unité académique : Département d'informatique et de recherche
opérationnelle, Faculté des arts et des sciences

Ce mémoire intitulé

Tactical Block Planning for Intermodal Rail Transportation

Preśenté par

Gianluca Morganti

A été évalué par un jury composé des personnes suivantes:
Bernard Gendron
Jean-Yves Potvin
Emma Frejinger
Teodor Gabriel Crainic

Mémoire présenté à la Faculté des études supérieures et postdoctorales en vue de l'obtention du grade de Maître dès sciences (M.Sc.) en Informatique

Résumé et mots-clés

Le mémoire présente le problème de la planification tactique des “blocks” pour le transport ferroviaire intermodal, qui a été peu étudié jusqu'à présent. Nous proposons un nouveau modèle de design de réseau en tenant compte de la spécificité du transport intermodal. La recherche se concentre sur le contexte nord-américain et fait suite à une étroite collaboration avec l'une des principales compagnies ferroviaires nord-américaines.

Le “blocking” constitue une importante opération de transport ferroviaire de marchandises, par laquelle des wagons d'origines et de destinations potentiellement différentes sont regroupés pour être déplacés et manipulés comme une seule unité, ce qui permet des économies d'échelle. La littérature se limite aux travaux traitant le problème classique du blocage des trains, où la demande est exprimée en termes de wagons. À notre connaissance, aucun travail préalable n'a été consacré à un contexte de transport intermodal, où la demande est exprimée en termes de conteneurs à déplacer d'un terminal d'origine donné vers un terminal de destination donné, introduisant ainsi un processus de consolidation supplémentaire.

Nous proposons un modèle de “blocking” qui prend en compte plusieurs types de conteneurs et wagons, intégrant l'affectation conteneur-wagon. Nous présentons un nouveau modèle de design de réseau à trois couches en temps continu formulé sous la forme d'un programme linéaire mixte en nombres entiers (MILP), dans le but de minimiser le coût total de transport composé par la sélection de blocs, les coûts d'exploitation et la gestion du coût de la demande. Le modèle peut être résolu en utilisant un solveur commercial pour des tailles réalistes. Nous illustrons les performances et l'intérêt de la méthode proposée à travers une étude de cas approfondie d'un important chemin de fer nord-américain.

Mots-clés: Rail intermodal; blocking; design de réseau; étude de cas.

Summary and keywords

The thesis presents the tactical block-planning problem for intermodal railroads, which has been little studied so far. We propose a new block service network design model considering the specificity of intermodal rail. The research focuses on the North American context and follows a close collaboration with one of the major North American railroad companies.

Blocking constitutes an important rail freight transport operation, by which cars with potentially different origins and destinations are grouped to be moved and handled as a single unit, yielding economies of scale. The literature is limited to works addressing the classical train blocking problem, where demand is given in terms of cars to be blocked among specific OD pairs. To the best of our knowledge, no prior work has been dedicated to an intermodal transportation context, where demand is expressed in terms of containers to be moved from a given origin terminal to a given destination terminal, hence introducing an additional consolidation process.

We propose a blocking model that considers several types of containers and railcars, integrating the container-to-car assignment. We present a new continuous-time, three-layer service network design model formulated as a Mixed Integer Linear Program (MILP), with the objective of minimizing the total transportation cost composed by block selection, operation costs, and handling demand cost. The model can be solved using commercial solver for realistic sizes. We illustrate the performance and interest of the proposed method through an extensive case study of a major North American railroad.

Keywords: Intermodal Rail; Blocking; Service network design model; Case study.

Contents

List of Figures

List of Tables

Acknowledgments

The thesis becomes reality with the help of many individuals to whom goes my sincere gratitude.

Foremost, I would like to thank my advisors Teodor Gabriel Crainic, Emma Frejinger and Nicoletta Ricciardi for the constant support and supervision throughout the years. Without your help it would not have been possible to complete successfully this long journey.

Also, I gratefully acknowledge the funding provided by the CN Chair in Optimization of Railway Operations, Université de Montréal, by the Natural Sciences and Engineering Council of Canada (NSERC), through its Cooperative Research, Discovery, and Acceleration Grants programs, and by the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) of Italy, through its Research Projects of Relevant National Interest (PRIN) program.

I also gratefully acknowledge the support of Fonds de recherche du Québec through their strategic infrastructure grants.

Last but not the least, an immense thank to my family for the endless support during these years.

Chapter 1

Introduction

This thesis presents the tactical block-planning problem for intermodal railroads, which has been little studied so far, by proposing a new block service network design model considering the specificity of intermodal rail. The research focuses within the North American context and follows a close collaboration with one of the major North American railroad companies.

The significant increase of demand for freight transportation in the last decades requires an always more efficient, cost-effective and environmentally sustainable freight transportation system. This includes the use of containerized, intermodal, transportation, which in today's world forms the backbone of world trade, and aims at integrating various modes and services of transportation to improve the efficiency of the whole process. The European Conference of Ministers of Transport defines intermodal transportation as "*the movement of goods in one and the same loading unit or vehicle, which uses successive, various modes of transportation without any handling of the goods themselves during transfers between modes*". This reduces cargo handling, which improves security, reduces damages and loss and allows a faster service at a lower cost.

Specifically, intermodal rail is the long-haul movement of shipping containers and truck trailers by rail, combined with a (usually shorter) truck movement at one or both ends. This way of transportation is characterized by the exigence of moving demand from an origin terminal to a destination terminal through the utilization of trains.

Railroads are at the heart of freight transportation systems, moving a broad variety of commodities long distances in a cost-effective and environmentally sustainable manner. In particular, they are a key element of the intermodal transportation network, which displays a steady traffic growth as illustrated by the the 5.3% annual increase (average, since 1990) at North American ports (IAPH 2015).

Efficient and profitable railroad activities require adequate planning of operations and resources. We focus on the *block and car tactical planning problem* that arises in intermodal rail transportation, using the operations of a major North American railroad as a case study. Blocking constitutes an important rail freight transportation operation. A block is a group of cars with possibly different origins and destination, which moves as a single unit between a pair of yards, without cars being handled individually (e.g., sorted) at intermediate yards. Blocking aims to take advantage of economies of scale and reduce the cost of handling cars at yards. A block is moved by a sequence of trains, while, in the general case, a car can be moved by one or a sequence of blocks between its origin and destination yards.

To the best of our knowledge, this work is the first one focusing on the intermodal rail block problem. There are, however, several studies focusing on the classical train blocking problem in the general railway system context. Then, the problem consists in assigning cars to blocks with given OD pairs, and demand is represented as number of cars to be blocked (Bodin et al. 1980, Newton et al. 1998, Barnhart et al. 2000, Ahuja et al. 2007). Blocking has also been addressed in the broader setting of tactical planning of rail services and schedule (Crainic et al. 1984, Crainic & Rousseau 1986, Zhu et al. 2014). In our context, the situation is different. Intermodal demand is defined in terms of containers of a certain type that necessitate to be transported from an origin to a destination, while it is tonnage translated into number of railcars of a particular type in the other case. About 90% of the containers used worldwide are either 20 or 40 feet, while longer units (45, 48 and 53 feet) are also used in North America market. This gives rise to a demand structure where shares of the different types of containers to transport vary over origin-destination (OD) pairs. Containers need to be assigned to railcars, and multiple containers can be placed in a single railcar. Hence, introducing an additional consolidation process. In terms of equipment, intermodal railcars are different from other general cargo ones. Cars with one to five platforms are used in North America with often the possibility to be double-stacked. Railcars have different loading capabilities depending on the number of platforms, their length and weight holding capacity. The multitude of railcar and container types results in a larger number of ways to load containers that must respect different loading rules (Mantovani et al. 2017).

The work of this thesis resulted in a publication: *Morganti, G., Crainic, T.G., Frejinger, E. and Ricciardi, N. (2019). Block planning for intermodal rail: Methodology and case study, Transporation Research Procedia 47: 19-26.* We summarize the contributions as follows:

- We propose a new block service network design model, using a three-layer time space network (container, block/car, train). It is a Mixed Integer Linear Programming formulation, with the objective of minimizing the total transportation cost composed by block selection, operation costs, and demand handling cost.
- We address challenges associated with the multitude of railcar and container types by adequate assignment constraints.
- We analyze performance results and provide managerial insights based on an extensive numerical study using realistic data.

The remainder of the thesis is structured as follows: Chapter 2, presents an overview of intermodal transportation and the tactical blocking problem for intermodal rail with its related literature. Chapter 3 is the research paper. Finally, Chapter 4 concludes.

Chapter 2

Problem Setting and Literature Review

2.1 Intermodal transportation

Nowadays, intermodal transportation forms the backbone of world trade. It refers to the transportation of people or freight from their origin to their destination by a sequence of at least two transportation modes, the transfer from one mode to another one is performed at an intermodal terminal (sea port, inland terminal, etc.).

Intermodal transportation demonstrates an increased complexity due to the use of multiple means of transportation and the involvement of multiple actors and stakeholders. This term implies integration between different operators in the transport chain, where these different transport modes should not be only optimized separately, but they should be attuned to one another. Indeed, contrary to conventional transportation systems in which the various transportation means work in an independent way, intermodal transportation integrates several transportation modes to enhance the performance of the whole transportation chain.

Intermodal transportation heavily relies on containerization. In fact, containers furnish several advantages that reduce cargo handling and accelerate operations both at yards and across the transport chain. A container, is defined as *"a box to carry freight, strong enough for repeated use, usually stackable and fitted with devices for transfer between modes"* by the European Conference of Ministers of Transport. Containers are large, uniform boxes used for transporting customer goods from one origin to a destination, and come with different characteristics, such as: size, length, type of goods transported.

An intermodal chain may be qualified by some consolidation transportation services, where different customers' loads move together in the same transportation mean, with possibly different origin-destination, or by some customized transportation systems, where carriers provide special specific services to each customer. Example of customized transportation modes are trucks, which move to customers location, where each truck is loaded and then moves to a specific terminal where it is unloaded. Thus, the service is tailored for each customer. Customized transportation is not always the most appropriate choice for shipping goods.

The trade-off among cost, frequency, delivery time of transport and the frequency and volume of shipping, often imposes the utilization of consolidation transportation services. Freight consolidation transportation is performed by Less-Than-Truckload motor carriers, railways, ocean shipping lines, etc. Consolidation transportation carriers and intermodal transportation systems are normally organized as hub-and-spoke networks (Fig.2.1). Observing the picture, services are offered among some origin and destination nodes, represented by numbered nodes. In order to exploit economies of scale, low volume of traffic travels until a consolidation point (called hub), which may represent seaport terminal, as well as rail yard, etc., where it is classified and consolidated to be moved in between hubs through high frequency and high capacity services. Traffic is successively relocated on lower frequency services to reach its final destination. A hub-and-spoke organization allows higher frequency of service among all origin-destination pairs in the network and a more efficient utilization of resources.

Terminals are a key component of consolidation and intermodal transportation systems, and their efficiency is vital for a high performance of the whole transport chain.

Several types of transportation modes are concerned with intermodal transportation:

- *Land transport modes:* Offered by LTL motor carriers and railways. On the one hand, truck transportation ensures flexibility, high frequency and low cost trasnportation. On the other hand, it leads to high congestion on roads, when the flow of trucks is high, as well as significant volumes of pollutants. Thus, when a large amount of

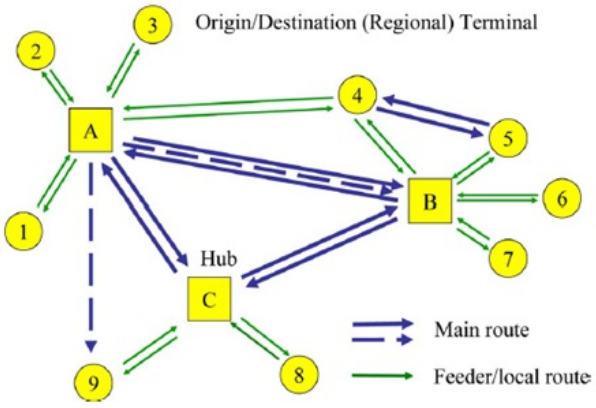

Figure 2.1: Hub and Spoke Network

containers needs to be transported, particularly over long distances, rail transportation is preferred, given its higher economic and environmental efficiency.

- *Air transportation mode:* Used for intercontinental intermodal transportation, but, due to the high fixed costs it is typically used for time sensitive, valuable or perishable freight carried over long distances.

- *Maritime mode:* It is similar to land and air modes, operates in its own space, which is at the same time geographical by its physical attributes, strategic by its control and commercial by its usage. Rivers and oceans are the two major elements that compose the physiography of maritime transportation. Although they are connected, each one represents a specific domain of maritime circulation. It is mostly container based, and its efficiency is given by the capacity of utilization of large container vessels, used for inter-continental travels. In maritime transportation, travel and loading/unloading time are larger compared that of land transport.

As part of the thesis, we are mainly interested in rail intermodal transportation, that is described in next section.

2.1.1 Intermodality through Railway Systems

Intermodal rail transportation typically constitutes the long-haul movement of shipping containers and truck trailers by rail, combined with a (usually shorter) truck movement at one or both ends. This means of transportation is characterized by the exigence of moving demand from an origin terminal to a destination terminal through the utilization of trains.

Intermodal terminals come in several designs and sizes, and may be specialized for particular transportation modes, offering special services, and the handling of specific products. In fact, each terminal has distinct characteristics: configuration, size, capacity, number of tracks. All those features may allow or not allow some important operations, such as classification, consolidation, etc. In the context of intermodal rail transportation, there are two different types of terminals: inland and port terminals.

Inland Terminals

An inland terminal is a special transshipment node in a rail network where loads for trains are processed (collected, rearranged, unloaded, stored, loaded, picked up, etc.). It can be a terminal itself or it can be located in (or nearby) a seaport.

An inland container terminal is commonly composed of the following elements, each performing a specific function.

- **Intermodal yard**. The core of the terminal where unit trains are loaded and unloaded by cranes (rubber-tired gantries) or lifts (side-loaders). They can be more than 2 km in length due to the large size of container unit trains. Containers are brought trackside or to the storage area by specific handling equipment. While older generations of intermodal yards (or those with small volume) worked on a one-to-one basis (one trackside space available for loading or unloading for each car) using reach stackers, more modern intermodal yards tend to operate on a two-to-one basis (one trackside space for loading and one trackside space for unloading). Terminals with high throughput are operated with rail-mounted gantry cranes able to straddle over several tracks (up to about 8) and are able to use track-side stack piles; therefore part of the storage area is within the intermodal yard.

- **Storage area**. Acts as a buffer between the road system (drayage) and the intermodal yard. It often covers an area of similar size as the intermodal yard since modern rail intermodal yards are heavy consumers of space. Storage in the intermodal yard can be grounded where containers are stored by stacking them upon one another, or wheeled with containers stored on chassis. In wheeled terminals, which are common in North America, containers are directly transferred to a chassis waiting to be picked up for delivery, and thus the chassis is an active element of terminal operations. On terminals that are more recent and thus have more space available, container/chassis pairs are parked in a parallel fashion to facilitate drayage to and from the intermodal yard. There is also some storage areas for refrigerated containers with power outlets, but this account for a small amount of the total storage capacity.

- **Classification yard**. Its function is mainly related to the assembly and break down of freight trains carrying other types of cargo. This is necessary because each rail car can be bound to a different destination and can be shunted on several occasions, which takes place at the origin, destination or at an intermediary location.

- **Gate**. This is where the truck driver presents proper documentation for pick up or delivery. Most of the inspection is done remotely with cameras and intercom systems where an operator can remotely see for instance the container number of an existing truck and verify if it corresponds to the bill of lading. To simplify matters and increase throughput, there are often separate entry and exit gates and dedicated lanes for empties or chassis returns.

- **Repair / maintenance**. Area where regular maintenance activities of the terminal's heavy equipment is performed.

Planning issues of rail terminals can be classified into categories according to the considered planning horizon. At the strategic level, the main problems regard the acquisition and/or construction of durable resources that should remain active over a long period of time. The tactical level generally involves the specification of operating policies that are updated every few months. The operational level mainly addresses problems occuring on a daily basis. Overviews of the related literature can be found in Carlo et al. (2014), Vis & De Koster (2003), Steenken et al. (2004), Macharis & Bontekoning (2004) and Boysen et al. (2013).

Demand and Rail Cars

Demand for intermodal transportation can be expressed in terms of number of containers with certain characteristics, encompassing size, weight, type of goods, origin and destination terminals, and time required at destination.

For the transport, containers are placed onto *rail cars*. The corresponding operational decision-making problem is called the load planning problem (LPP) introduce by Mantovani et al. (2017). In the following we briefly describe rail car characteristics which are crucial to the LPP.

Rail cars are of different size and type, and based on their features (car design, number of platforms, length of the platform, capacity, stacking configuration), they may allow different loading configurations. The latter express both the number of containers they can carry, and the way of loading. Most car types in the North American fleet allow the double-stack of containers, which means that containers are allowed to be placed one on top of another one in the same platform of the vehicle. Other cars, such as spine cars, allow instead only single stacked containers. Each container is loaded in one of the so-called *slots* of the *platforms*, which corresponds to the physical spot for positioning containers on the car. Slots can be bottom or top slots, and may have different lengths. More than one container can be placed in a same slot (two 20-foot containers in a bottom slot). Considering all the different characteristics, some cars are then more appropriate for the loading of specific containers. In fact, some of those may require specific cars to be moved around the network.

The loading of containers onto rail cars must respect a number of physical constraints. Each type of rail car has a set of feasible configurations, which need to be observed in order not to violate specific loading rules. Furthermore, dimensional restrictions on loading a set of containers on a set of available cars need to be considered. In fact, each platform has an allowable weight holding capacity, and the total weight of the loaded containers cannot exceed it. For double-stack platforms, the center of gravity must be respected. Last but not least, other technical constraints concerning the way of loading some types of containers must be respected.

Rail cars are consolidated into so-called *blocks* when moved by a train. The railcars in a given block are not reshuffled until the final destination of the block is reached. Rail cars assigned to a train are usually contiguously grouped so as to form blocks, in such a way that the total block distance and the number of re-classifications is minimized. However, it is not possible to assign all demands to one block, since blocks have a maximum weight, maximum length and maximum number of cars that can be assigned to. Each block travels from an origin, where it is assembled, to a destination, where it is disassembled, and demand is consequently re-classified into new blocks, and moved further away. Formation of blocks takes place within terminals, where classification tracks make it possible to properly sort and assign the rail cars.

Once blocks are determined, they are grouped together to form intermodal trains, respecting the schedule and capacity of the trains.

Capacity

The operations of building up trains require enough capacity at each yard where the procedure is performed, as well as specific equipment available for making the operations possible. Capacity is an important factor to be considered, particularly within inland terminals, where trains need to be built up. Thus, the number of classification tracks has to be sufficient for allowing cars to be sorted and consolidated, as well as the possibility of carrying out some operations such as blocks swap, block composition, and block disassembly.

2.2 Blocking Problem

The fundamental concept of aggregating freight railcars, based on various attributes, to create blocks and subsequently of combining blocks to create trains goes back to the early history of the freight railroad industry. In the general cargo case, railroads receive requests from customers to transport cars. Upon receiving the request, based on each car's attributes (such as physical dimensions, freight type, etc.), the railway generates a trip plan detailing the movement of the car from the customer's origin location to the requisite final destination.

Since each major railroad ships every year millions of cars from their origins to their destinations over their rail networks, the movement of shipments has to be dictated by an operating plan that makes it possible to efficiently move demand within the network. Current industry approach involves the resolution of two main issues:

- Blocks to build up at each rail yard (Railcar-to-Block assignment)
- Trains to offer (including Block-To-Train Assignment, Train Routing and Train Scheduling)

Train routing is concerned with the identification of the origin, destination and route for each individual train, so that these paths are consistent with the rail network and the blocks to be transported. During its trip, a train might visit different yards to perform work events like pickup blocks, remove blocks, or both operations. Depending on the work event, the train is required to stop for a certain amount of time. Thereby, train routing also includes determining schedules of the train, consisting in defining arrival and departure time from each yard, as well as some specific features of the train (speed, power, capacity, etc.). The plan is created from few weeks up to few months in advance, before being executed, and the plan is then followed and adjusted (when necessary) throughout the period.

2.2.1 Traditional Blocking

The blocking problem arises at the tactical level. It is fundamentally a consolidation problem, since it addresses the problem of how to consolidate a large number of shipments in a way to decrease the necessity to handle them during their trip from their origin to their destination. Reducing handling operations lead to a minimization of delays and total transportation costs. The problem arises from the necessity of reducing time delays caused by the reclassification of the cars at intermediary yards on their journeys. Thus, railroads group shipments instead, building the so-called blocks. A block is a set of rail cars, with an associated OD (origin-destination) pair, which most of the time does not necessarily match with the OD of any cars composing it. The advantage of forming blocks is that shipments placed on it cannot be touched until the block has reached its destination in the rail network. Ideally a shipment would be allocated to a direct block, where origin and destination match with those of the block, in order to avoid re-classification and delays. Constructing a block requires a certain number of operations and a certain amount of yard capacity. Yard resources and capacities limit the maximum number of blocks and cars a specific terminal can handle, thus rejecting the idea of assigning direct blocks for each individual shipment. Owing to the consolidation process, some shipments may be forced to travel longer distances with respect to their shortest path in case of a direct route from their respective origins to their destinations. Furthermore, they may go through the process of intermediate handling and switch trains.

Railroads seek to develop a block plan ruling which blocks should be built at each yard and which shipments should be allocated in every block. The plan attempts to move all the traffic demand by trying to minimize the number of classifications and the final transportation costs around the network. At the network level, companies construct a cyclic plan for a fixed period. As that plan is tactical, it is updated only few times over the whole year, and once it has been constructed it is repeated.

Literature Review

The blocking problem has mostly been addressed as a separate problem, without considering other related problems, e.g., the scheduling of train services. One of the first contributions was proposed by Bodin et al. (1980) that provides a multi-commodity flow formulation to solve the blocking problem. Demand is defined in terms of cars that have to be moved from one yard to another in the network. It aims to determine the optimal blocking strategy for all the yards simultaneously. They use a piece-wise linear objective function to minimize delay costs at each yard. They consider block capacity constraints at each yard, impose the length of a block to be between a lower and an upper bound of the number of cars, and pure strategy constraints to ensure that all commodities traveling between two terminals follow the same blocking path.

Newton et al. (1998), modeled the blocking problem as a network design problem. They present a mixed integer linear programming (MILP) formulation. They study the problem without explicitly considering the movement of the trains, so they do not account for the assignment of blocks to trains. They assume a train schedule to be developed after the blocking decision. They also assume no block swaps performed within terminals. Demand is expressed in terms of cars that have to go from one origin to a destination. It is subdivided into priority classes, which basically consists in dividing shipment between express and nonexpress, according to their priority. The objective function seeks to minimize the number of handling required to deliver commodities, which implicitly limits the congestion effects in each yard and their overall usage. The model places limit on the maximum number of blocks that can be built at yards, and the maximum number of cars re-classifications at each yard. Moreover, they impose a maximum number of re-classification for each shipment class. They use a column generation, branch-and-bound algorithm in order to generate attractive blocking paths, and reduce the number of design variables before solving the model.

Barnhart et al. (2000), propose a network design model similar to the one of Newton et al. (1998), but they utilize a dual-based Lagrangian relaxation approach to solve it, which decomposes the problem into two simple sub-problems. They do not account for train movements, and they suppose that the train schedule is constructed after having solved the blocking problem. Demand is given in terms of cars with a specific origin-destination pair. They minimize the number of handling, assuming handling costs and delays proportional to the number of handling. The solution is limited by the usual terminal constraints, maximum number of cars that can be classified at the yard, and maximum number of blocks that can be built at the yard. They do not associate fixed costs for the construction of the block, they use forcing constraints to divide the problem into two easy sub-problems: one multi-commodity flow problem, and one integer block formulation.

The aforementioned studies focus on the blocking problem in isolation from other related decision-making problems. Crainic et al. (1984) proposed one of the first models addressing at the same time the selection of services and their frequencies, the classification and the blocking of the cars, the makeup of trains and the freight routing. Demand is considered in terms of cars to be moved among two different yards, and it is separated into different classes, according to the costs of handling different types of goods. They use a non-linear objective function that takes into account trains through delay costs, which discourage small blocks to be built at yard. Itinerary vectors that specify which block and train carry each commodity are introduced through column generation. They use an algorithm that works on a modified version of the problem, in which the capacity constraints are integrated in the objective function, and not explicitly as model constraints. They solve the problem via a decomposition method approach.

Zhu et al. (2014) proposed the first modeling framework for key tactical planning decisions for freight rail transportation, by integrating service selection and scheduling, classification and blocking of cars, train routing and makeup. They use a three-layer space time network representation of decisions, operations and their dependency, where layers are services, blocks and cars. The model (MIP) encompasses the service selection, block selection and car distribution decisions, limiting the car classification and the block handling in each yard at each period, as well as the running of multiple trains on tracks at each period. The objective is to minimize the cost of the system during the schedule duration, where the cost is given by multiple factors. They propose a new meta-heuristic solution method that combines slope scaling and ellipsoidal search.

2.2.2 Blocking of Intermodal Traffic

In the intermodal transportation context, as opposed to the traditional blocking problem, demand is defined demand defined in terms of the number and features of the containers that need to be transported from an origin terminal to a destination terminal. The definition of the demand makes the problem fairly different from the one described in the previous section. Indeed, in the traditional blocking problem one rail car corresponds to one unit of demand, while in the intermodal case, this is no longer true since one railcar can carry more than one container, sometimes as many as 15 containers.

The peculiarity of the intermodal blocking problem is that, instead of only blocking railcars into blocks, to minimize the total transportation costs and the number of re-classifications, we face three consolidation processes: assigning of demand (containers) to rail cars, forming blocks of railcars and assigning blocks to trains.

We recall that containers differ depending on different characteristics. In North America they have different lengths (20, 40, 45, 48, 53 feet), and they transport different types of cargo. A container can be classified according to the following main types: dry, refrigerated, liquid, dangerous. In addition, a container may be classified depending on its loading status, which is either full or empty.

In the transportation process, containers are also characterized by their origin-destination pair, which indicates the terminal where they start and end their journey. Containers might also be required to be at a certain destination within an established time (due date).

Demand has to be transported over the rail network and carried to a defined destination via the utilization of rail cars. The latter are characterized by the following features:

- *Number of platforms*: 1, 3, 4, 5, 7
- *Platform length:* 40, 45, 48, 53, 56 feet
- *Weight limits:* exact or qualitative depending on specified thresholds.

The way of loading containers depends on the possible feasible loading configurations of the rail cars. In fact, according to their aforementioned characteristics, some rail cars may allow the loading of only some specific types of containers rather than others. Furthermore, we may have rail cars allowing containers to be loaded one on top of another (double stack configuration), according to specified rules and constraints, or not allowing it (single stack configuration).

2.2.3 Container Operations

According to some company policies rules, once containers are loaded onto rail cars in its origin terminal (where it has arrived by truck, train or vessel), two possible operations may occur: reloading or no reloading of containers. Depending on which policy is chosen, different operations and different way of loading may be performed.

1. *No reloading operations:* Not allowing a container to be unloaded from a rail car during its trip implies that the containers stay on the same car until their destination in the rail network is reached. Once they arrive at their destination, they are off loaded and transported further towards the customer destination. Hence, containers positioned on the same car have the same rail network destination and we can block containers and rail cars together. This leads to an important implication, namely the transformation of the demand, from the number of containers of a certain type going from an origin towards a destination, to the number of loaded cars going to a certain destination.
2. *Reloading operations:* Reloading means that at least once before reaching its final destination a container can be unloaded from a rail car and loaded onto another one.

On the one hand, the extra planning and handling operations involved in reloading take place at a cost. On the other hand, reloading affords additional flexibility, as for instance, it can allow the utilization of empty slots (available places for loading containers on cars) that would otherwise be vacant.

The decision to reload or not taken by each company must be the result of a carefully considered trade-off between the costs of leaving empty slots and the additional costs incurred in planning, synchronization and handling.

2.2.4 Car Availability

Each railroad owns or leases a fleet of rail cars. As mentioned in the previous section, rail cars come in different sizes and types. Obviously, the number of resources available for transporting demand is limited at each yard, meaning that a problem of managing the fleet occurs in solving the blocking problem. The fleet can be assumed to be homogeneous or encompassing several types of rail cars, that may partially, totally or not at all substitute for one another. This fleet needs to be managed concurrently, figuring out how to move resources inside the network, in order to operate in an efficient and effective way the network, by trying to have always available specific resources for a specific type of demand, when requested. Hence, the management of resources can be faced in two different ways:

1. *No resource management:* Neglecting the management of resources implies that rail cars assumed to be always available when they are necessary. Thus, at each period, in each terminal the best resources required are there, and are used for the demand transportation. The solution to a problem without resource management is an optimistic one and hence can provide a useful bound.

2. *Resource management:* When we consider a limited fleet of cars, the situation radically changes. Indeed, how a resource is moved within the network, and for which type of demand it is used, needs to be addressed as part of the block plan. We may have two different cases, when the fleet of resources is managed: We exactly know the available cars at each terminal, or we just know the total fleet at disposal of the company. In the former case, an initial distribution of rail cars in the network is given, assuming

resources available at each terminal during the schedule length considered. In the latter, instead, the distribution of cars inside the network is unknown, thus, we do not have any information about the precise location of the resources, but we may have an external node from where resources are spread in the network where necessary.

This thesis focuses on the blocking problem without resource management.

2.3 Problem Statement

The problem we address is concerned with the determination of the tactical blocking plan given a demand structure, thus, we identify the optimal assignment of rail cars to blocks and blocks to trains in order to satisfy the demand for the transportation of containers. The blocking plan is developed for a schedule length (e.g., a week) and is meant to be applied throughout a season or possibly over a longer period according to the railroad company policies and to the perceived shifts in the business and network environments. Its construction is data-driven: it is based on forecasts of the demand for the transportation of containers inferred from the available historical data.

The particularity of our problem lies in the fact that the development of an efficient blocking plan is conditioned by a given railroad network, since train schedules, train length and train capacities are known in advance. Assuming the schedule as known and fixed, provides information about all the routes of all the trains, and their timetables, as well as their main features. Hence, for each train we precisely know the origin, the destination and the intermediary stops, as well as the arrival and departure times from each yard, and their capacity, power, length, and some other characteristics.

We represent the problem with a multi-layer time-space network whose time discretization is induced by the train schedule. As mentioned above, for each train we exactly know its timetable from and to each yard in the network. Blocks have to be built or transferred a certain time in advance, to be assigned to a specific train passing through that yard at a specific time.

The knowledge of the train lengths (capacity) limits the size of the blocks in addition to the capacity constraints at each terminal. A block may be viewed as a path followed over the network, to which is assigned a set of rail cars with a loaded demand.

As mentioned in Section 2.2.2, in an intermodal context, the blocking problem faces two consolidation processes. The major difficulty we face comes from the initial loading of the demand into a set of rail cars. Given a set of containers of different types in each period of the schedule length, we seek to define the best (approximate, since we face a tactical level problem where a deep level of detail is not required) assignment of demand to rail cars, in order to satisfy as much demand as we can. Simultaneously we need to respect the handling constraints at terminal, which depending on the number of classification tracks in the yard, determine the maximum number of rail cars that can be classified.

In this research, we initially assume that there is no reloading operations en-route. This assumption implies that all containers loaded onto the same rail car must have the same destination. In this manner, it is possible to transform demand into number of loaded rail cars that travel from an origin towards a determined destination. Moreover we do not consider a limited railcar fleet and hence do not manage resources.

The ultimate goal of this research is to develop an optimization model that can be a component of a data-driven decision-aid-tool. Such a tool can help decision makers at the tactical planning level to make better and well-informed decisions which, in turn, may lead to reduced cost and increased demand satisfaction.

Chapter 3

Block planning for intermodal rail: Methodology and case study

Gianluca Morganti Teodor Gabriel Crainic Emma Frejinger Nicoletta Ricciardi

Author contribution

I designed and implemented the first version of the three-layer network representation and the corresponding MILP model. I generated the first set of results. Moreover, I produced the first draft of the text for the article.

3.1 Introduction

Railroads are core elements of intermodal transportation, moving containers loaded with a broad variety of commodities over long distances in a cost-effective and environmentally sustainable manner. Efficient and profitable railroads require adequate planning of operations and resources. We focus on the tactical block-planning problem arising in intermodal rail transportation, which has been little, if at all, addressed until now.

A block is a group of cars with possibly different origins and destinations, which moves as a single unit between a pair of yards, without cars being handled individually at intermediate yards. Blocking aims to take advantage of economies of scale and reduce the cost of handling cars. A block is moved by a sequence of trains, while a car can be moved by one or a sequence of blocks between its origin and destination (OD). The classical train blocking problem consists in selecting the blocks to build and assigning cars with given OD pairs to blocks (e.g., Ahuja et al. 2007, Ahuja & Şahin 2014, Jha et al. 2008, Newton et al. 1998, Barnhart et al. 2000, Bodin et al. 1980). Blocking has also been addressed in the larger setting of tactically planning the service and schedule network of railroads (e.g., Crainic et al. 1984, Crainic & Rousseau 1986, Zhu et al. 2014, Crainic & Kim 2007). Existing methods assume, however, that the demand of the blocking model is given as a number of cars to be blocked. Consequently, they do not account for the important characteristics particular to intermodal rail, the most obvious one being demand expressed in numbers of containers of different types rather than in numbers of loaded cars. This means that container assignment to and loading on cars must be explicitly included in the blocking model. The multitude of car and container types results in many ways to load containers that must respect different loading rules, particularly when double stacking is allowed. Representing in a computationally efficient way the assignment of containers to cars within a tactical blocking model is particularly challenging. Our research objective is to contribute filling these gaps in knowledge and decision-support instruments, and propose a blocking model that considers several types of containers and cars and integrates the container-to-car assignment.

Section 3.2 describes the tactical block-planning problem for intermodal rail. The modelling framework is presented in Section 3.3. Section 3.4 is dedicated to the numerical experiments performed, and the presentation of the performance results and associated managerial insights. Finally, Section 3.5 concludes.

3.2 Problem setting

Railroads operate according to two levels of consolidation. First, cars are grouped together within blocks to be moved together, as a unique entity, from the origin of the block, the terminal where it is constructed, to its destination terminal, where it is dismantled. The cars making up a block do not necessarily have the same origins or destinations, and these are not necessarily the same as those of the block they are grouped into. They share, however, a part of their respective trips. The second consolidation level puts together blocks to make up trains. Consequently, a block travels from its origin to its destination either on a unique train or on a sequence of trains. In the latter case, the block is transferred as a unit between two consecutive trains at an intermediate terminal. Trains are made up either of a single block or a sequence of blocks. A train is made up at its origin terminal and dismantled at its destination terminal, where its blocks are either transferred to other trains or dismantled as well. The route of a train gives its origin, destination, and intermediate terminals where it stops to drop or pick up blocks, while its schedule gives the departure and arrival times at all the terminals, origin, destination, intermediate, on its route.

The trains and blocks make up a service network in space and time aimed at moving the cars holding the freight from the origin to the destination (OD) of the respective demand. Cars are of different types. In the general, non-intermodal, case, the selection of the car types assigned to each customer is part of the commercial transaction between the railroad and its customer. Consequently, the assignment decision, as well as the activities related to the loading of the cars, is outside the scope of the blocking problem. The latter is thus concerned with cars only, not the freight they carry. Cars carrying freight for a particular OD travel on a sequence of blocks and trains. At the origin terminal, cars undergo classification, a sorting operation to put them to the track where the block to which they are assigned is being built. When the block arrives at its destination and is dismantled, the cars it contained are either delivered to customers, if at their own destination (how this operation is performed is beyond the scope of this paper) or, in the case of non-intermodal traffic, are directed to the classification part of the terminal to be sorted and attached to their new block. The blocking problem aims to select the blocks to build, including the train sequence, and the classification strategy, that is the assignment of cars to blocks. It belongs, together with the selection of train services and resource (e.g., locomotives and cars) management strategies, to the tactical planning set of decisions, yielding the operation plan for the next “season” (Crainic & Kim 2007). It also belongs to shorter-term planning decisions (e.g., the week) when plans need to be adjusted due to, e.g., incidents and unforeseen delays or demand variations.

Intermodal rail transportation follows the same general canvas, with a number of important characteristics that sets it apart and requires particular methodological developments. Reclassification of cars at intermediary terminals is seldom performed to avoid additional delays. Indeed, it is not included in the blocking plan of the major North American railroad we collaborate with. The most important difference, however, concerns the need to explicitly consider the selection of the type and number of cars assigned to an OD demand, given the particular number and type of the corresponding containers. Intermodal rail transportation thus implies a third consolidation operation of containers on multi-platform cars, which makes the corresponding blocking problem different and much harder than for regular traffic.

There is a large variety of container types (e.g., 40- and 53-feet long) and railroads use fleets of cars of various types, each with one or several platforms and slots on the platforms. Single- and double-stack platforms have one and two slots, respectively. The matching/loading of containers to slots given the car type is an important but complex issue since 1) not all combinations are legal or suitable, and 2) a very large number of loading alternatives exist (Mantovani et al. 2017). How to address the differentiation of car and container types and the large number of loading alternatives in a way appropriate for tactical planning is a challenge that we address in this paper.

We conclude this section by emphasizing that in the blocking literature referred to above, block planning is performed first, the train make up and selection coming after. The problem we address is different. Trains and their schedules are selected first, in a separate planning phase, and are part of the input to the blocking problem. The model presented in the following section addresses this problem setting. We also analyse a number of problem variants reflecting practice, e.g., the existence of a list of preferred blocks for each train and the possibility to not move the complete demand by paying a penalty cost or to split the demand among several blocks.

3.3 Modelling

The railroad operates over an infrastructure network, where a number of terminals $\theta \in \Theta$ are dedicated, totally or partially, to the intermodal traffic. The plan is built for a *cyclic* *schedule* of given *length* (e.g., a week), assumed to be repeated over the tactical planning horizon. A set of train services $\Sigma = \{\sigma\}$ is given with their respective origin, destination and intermediary-stop terminals, and schedules. We propose a *block service network design model* based on a *three-layer cyclic space-time network* $\mathcal{G} = \{\mathcal{N}, \mathcal{A}\}$ (Zhu et al. 2014). Different from most of service network design literature (Crainic & Kim 2007) we do not discretize time. We rather use a continuous time representation where time moments correspond to the arrival/departure times of services at terminals. Demand is defined in this context by its origin terminal, time at the origin terminal, destination terminal, due-date at destination, container type, number of containers of that type, and a penalty cost for late arrival. The container types in this version of the problem are the 40- and 53-feet container boxes; 20-feet containers are represented by transforming the corresponding demand into 40-feet containers; similarly, 45 and 48-feet containers are treated as 53-feet ones due to their loading mode. This hypothesis simplifies the presentation of the container-to-railcar assignment without impact on the generality of the formulation.

The network is schematically illustrated in Figure 3.1, for a terminal and time moments corresponding to two trains, one being at one of its intermediary stop, the other being at its origin. We first describe briefly each layer, nodes, and arcs, and then present the model formulation.

The known arrival and departure times of each scheduled train at the terminals on its route yield the corresponding arrival and departure nodes, TIN and TOUT in the *train layer*, defining the time structure of the entire network. We model train activities through *TrainHandling* arcs representing the time trains spend at terminals and *TrainMoving* arcs between terminals. The train features, e.g., maximum length, provide the capacity u_a of the moving arcs.

The *block (& car) layer* is illustrated below the train layer. Let $\mathcal{B} = \{b\}$ represent the set of *potential blocks* generated by the set of trains. A *block* is defined by its unique route and schedule from its origin to its destination, made up of movements on TrainMoving arcs and activities at terminals. We model building new blocks and assigning them to specific trains, transferring blocks from one train to another, and dismantling blocks at their destinations through the *Block2Train*, *BlockTransfer*, and *Block2Dismantle* intra-layer arcs, respectively.

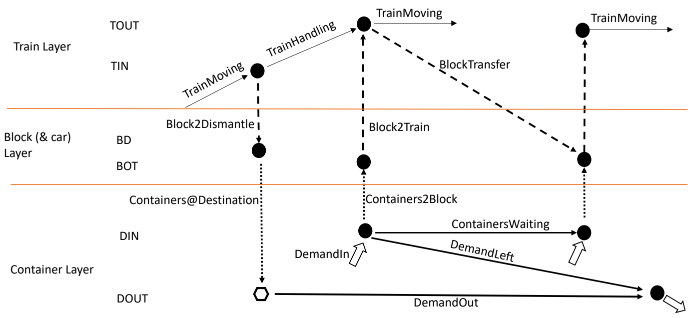

Figure 3.1: Three-Layer Time-Space Network Representation

Each TIN node thus generates a block-drop-off node BD in the block layer. Similarly, each TOUT node generates a BOT node in the block layer, where the new blocks just formed and the blocks transferred from arriving trains, all sharing the same first/next train on their respective routes, are collected before being attached to the train. Each block is thus characterized by its: *route* and *schedule*, given by the set of trains moving it from its origin terminal o_b to its destination d_b, through (possibly) a sequence of intermediary terminals; *capacity* u_b in terms of total block length; and *fixed cost* f_b, representing the total cost of building it and transferring at intermediate terminals, as well as costs for the total transfer and idle times (i.e., the train carrying it stops at a terminal for activities that do not involve it).

Demand enters and exits the system through the *container layer*. Demand $k \in \mathcal{K}$ is defined by its *origin terminal* o_k, *time at* the *origin* terminal α_k, *destination terminal* d_k, *due-date* at destination β_k, *container type* $\tau_k \in \mathcal{T}$, and *volume*, i.e., number of containers of that type, ν_k.

Let $\mathcal{B}_k$ be the set of *possible blocks* which may be used to transport demand k. A unit *transportation cost* c_{bk} is associated to each block $b \in \mathcal{B}_k$. The particular items included in c_{bk} vary according to the particular application. To better connect the selection of the blocks and the objectives of the railroad for a high-level of in-movement equipment (i.e., not idle), we include in the unit transportation costs c_{bk} measures related to the cost of hauling the cars, the train-to-train transfer time, the total number of transfers, and the idle time.

Upon arrival, containers wait on *ContainersWaiting* arcs until the selected block and departure time. The inter-layer *Container2Block* arcs support the container-to-car-to-block assignment model we propose, which is based on the relations among the lengths of blocks, platforms, and the two main container types (40 and 53 feet), as described below. The *Containers@Destination* inter-layer arcs support the flow of containers at destination, unloaded from cars out of the dismantled blocks. This flow then moves on the *DemandOut* arcs, paying possibly late-arrival penalty costs p_{bk}^{LATE}, until a super-sink node.

Two variants of the problem setting are of interest in relation to demand. First, whether the containers of a particular demand may be separated and delivered by several itineraries, using different blocks (and trains), possibly at different times. When such *splitting* is allowed, one may still wish to limit the number of selected blocks, e.g., through penalty-like itinerary cost p_k^{SPLIT}. The second variant concerns the case when the total train capacity is less than the total demand. In such cases, an artificial arc, indicated as *DemandLeft* in Figure 3.1, is added between the arrival node of each demand and the sink node, with a penalty-like unit cost of p_k^{LEFT} for the flow that cannot be delivered by the given rail network.

Recall that, different from most other commodities hauled by railroads, intermodal demand must be loaded onto cars at the origin terminal and must be unloaded at destination. We now describe the approach we propose to integrate the container-to-car assignment into the tactical-planning model. Recall that *length* is a major constraining feature for trains, as well as for blocks both in terminals and when put on trains, and that it is the capacity measure used in this model. Furthermore, cars come in various multi-platform configurations, a platform providing two slots, one on the bottom and one on top, for containers of a given length, 40 and 53 feet in our case. The length of a car is then determined for the most part by its number of platforms and, consequently, so is the length of the block. We therefore use the *platform*, of a given type, as our loading unit making up the block.

Let $\Gamma = \{\gamma\}$ be the set of platform types. Consistent with the container types, two platform types are considered in this paper, the 40- and 53-feet long (λ^{p40} and λ^{p53} represent their respective lengths). The basic loading rules considering these container and platform types are:

- *40-feet platform*: 1) single 40-feet container in bottom slot; 2) 2 40-feet containers in the two slots; 3) 1 40-feet container in bottom slot and a 53-feet one in the top slot;
- *53-feet platform*: 1) single 53-feet container in bottom slot; 2) 2 53-feet containers in

the two slots.

Because the overall goal of the loading problem is to maximize the number of forty-foot equivalent units per unit of train length, it is always best to use as many 40-feet platforms as possible. Under this hypothesis, when nb_{53}, the number of 53-feet containers, is greater than or equal to nb_{40}, the number 40-feet containers, then the 40-feet containers should be placed in bottom slots and the 53-feet ones on top, as much as possible. The number of 53-feet platforms, nbp_{53}, is hence given by Equation (3.1), which easily yields the number of 40-feet platforms, nbp_{40}, through Equation (3.2)

$$nbp_{53} = \max\left\{0, \lceil (nb_{53} - nb_{40})/2 \rceil\right\}, \tag{3.1}$$

$$nbp_{40} = \lceil (nb_{53} + nb_{40})/2) \rceil - nbp_{53} \tag{3.2}$$

We propose a mixed integer linear programming (MILP) formulation with decision variables:

- $y_b = 1$, if block $b \in \mathcal{B}$ is selected, 0 otherwise;
- $y_{bk} = 1$, if block $b \in \mathcal{B}_k,\ k \in \mathcal{K}$, is selected, 0 otherwise;
- x_{ak}, volume of demand $k \in \mathcal{K}$ on arc $a \in \mathcal{A}$, with x_{bk}, the volume of demand k on block $b \in \mathcal{B}_k$;
- x_k, volume of demand k left at the origin terminal, i.e., the volume on the corresponding artificial arc;
- x_b^{c40} and x_b^{c53}: number of 40- and 53-feet containers, respectively, assigned to block $b \in \mathcal{B}$;
- $x_b^{\pi 40}$ and $x_b^{\pi 53}$ number of 40- and 53-feet foot platforms in block $b \in \mathcal{B}$, respectively.

The integer programming formulation for the general problem setting, where the total train capacity might not be sufficient and the split of demand is permitted, may then be written as

$$\text{Minimize} \sum_{b\in\mathcal{B}} f_b y_b + \sum_{k\in\mathcal{K}} \left(\sum_{b\in\mathcal{B}_k} (c_{rk} + p_{rk}^{\text{LATE}}) x_{bk} + \sum_{b\in\mathcal{B}_k} p_k^{\text{SPLIT}} (y_{bk} - 1) \right) + p_k^{\text{LEFT}} x_k \tag{3.3}$$

subject to the following constraints

$$\sum_{b\in\mathcal{B}_k} x_{bk} + x_{ak} = \nu_k,\ k\in\mathcal{K}, \tag{3.4}$$

$$\sum_{b\in\mathcal{B}_k} y_{bk} \geq 1,\ k\in\mathcal{K}, \tag{3.5}$$

$$x_{bk} \leq y_{bk} u_{b_r},\ b\in\mathcal{B}_k,\ k\in\mathcal{K}, \tag{3.6}$$

$$x_{a_i\in\mathcal{A}_\theta^{\text{DIN}}k} + x_{a_i\in\mathcal{A}_\theta^{\text{DWT}}k} + x_{ak} = \nu_k,\ \ \theta = o_k,\ i : t(i) = \alpha_k,\ k\in\mathcal{K}, \tag{3.7}$$

$$x_{a_i\in\mathcal{A}_\theta^{\text{DIN}}k} = \sum_{b\in\mathcal{B}_{ki}} x_{bk},\ \theta = o_k,\ i\in\mathcal{N}_\theta^{\text{DIN}},\ i : t(i) = \alpha_k,\ k\in\mathcal{K}, \tag{3.8}$$

$$x_{a_{i-1}\in\mathcal{A}_\theta^{\text{DWT}}k} + x_{a_i\in\mathcal{A}_\theta^{\text{DIN}}k} = x_{a_i\in\mathcal{A}_\theta^{\text{DWT}}k},\ \ \theta = o_k,\ \forall i : t(i) > \alpha_k,\ k\in\mathcal{K}, \tag{3.9}$$

$$x_{a_i\in\mathcal{A}_\theta^{\text{DIN}}k} = \sum_{b\in\mathcal{B}_{ki}} x_{bk},\ \theta = o_k,\ i\in\mathcal{N}_\theta^{\text{DIN}},\ \forall i : t(i) > \alpha_k,,\ k\in\mathcal{K}, \tag{3.10}$$

$$x_{a_i\in\mathcal{A}^{\text{D2D}}k} = \sum_{b\in\mathcal{B}_k:\beta_b = t(i)} x_{bk},\ \ \theta = d_k,\ i\in\mathcal{N}_\theta^{\text{CD}},\ k\in\mathcal{K} \tag{3.11}$$

$$\sum_{i\in\mathcal{N}_\theta^{\text{CD}}} x_{a_i\in\mathcal{A}^{\text{D2D}}k} + x_k = \nu_k,\ \ \theta = d_k,\ k\in\mathcal{K}, \tag{3.12}$$

$$x_b^{\pi 53} = \max \left[0, \left\lceil \frac{1}{2} \left(\sum_{k:\tau_k = p53} x_{bk} - \sum_{k:\tau_k = p40} x_{bk} \right) \right\rceil \right],\ b\in\mathcal{B}, \tag{3.13}$$

$$x_b^{\pi 40} = \left\lceil \frac{1}{2} \left(\sum_k x_{bk} \right) \right\rceil - x_b^{\pi 53},\ b\in\mathcal{B}, \tag{3.14}$$

$$\lambda_b = \lambda^{p40} \sum_{k:b\in\mathcal{B}_k} x_b^{\pi 40} + \lambda^{p53} \sum_{k:b\in\mathcal{B}_k} x_b^{\pi 53},\ b\in\mathcal{B}, \tag{3.15}$$

$$\lambda_b \leq u_b,\ b\in\mathcal{B}, \tag{3.16}$$

$$\sum_{b\in\mathcal{B}_a} \lambda_b \leq u_a\ \forall a\in\mathcal{A}_\sigma^{\text{TM}},\ \sigma\in\Sigma, \tag{3.17}$$

$$y_b, y_{bk} \in \{0, 1\}, x_{bk}, x_{ak}, x_k \leq 0, \;\; k \in K, \;\; r \in \mathcal{B}_k, \;\; b \in \mathcal{B}, \;\; a \in \mathcal{A}. \tag{3.18}$$

The objective function (3.3) minimizes the total cost composed of the block selection and operation costs, plus the costs associated to handling demand: costs for moving and waiting, penalty for late delivery, splitting costs, and penalty for not delivering the full volume. Recall that the loading procedure automatically enforces the maximum utilization of the most desired platforms.

Constraints (3.4) ensure that all the volume of each demand is shipped out either on itineraries or on the artificial arc, while Constraints (3.5) specify that at least one itinerary must be selected for each demand. The latter constraints enforce equality to 1 when demand splitting is not allowed. The linking constraints (3.6) enforce the rule that flow is shipped out only on selected itineraries and that it is not larger than the capacity of the corresponding block.

Flow conservation at the origin of the demand is taken care of by Constraints (3.7) - (3.8), the latter link the number of containers to be sent at that time to the blocks they are assigned to. The next two, (3.9) - (3.10), perform the same role at the subsequent DIN time moments. Constraints (3.11) - (3.12) complete the flow-conservation task at the destination of demand. Constraints (3.11) link the flow of commodity k on the exiting arc to volumes carried by the blocks that brought it to the destination terminal. Constraints (3.12) then make sure all containers are delivered at destination.

Constraints (3.13) - (3.14) compute for each potential block the number of loaded platforms of each type corresponding to the commodity traffic assigned to them (Equations (3.1) - (3.2)). Constraints (3.15) - (3.16) compute the block length and enforce the block capacity constraint, while Constraints (3.17) enforce the train capacity on each of the respective moving links. Constraints (3.18) define the feasible domain of the formulation.

3.4 Experimental Results & Analyses

Our objectives for the experimental campaign were to 1) evaluate the efficiency of addressing the blocking model for realistic settings using off-the-shelf commercial software; 2) study the impact of a number of important problem characteristics on the blocking plan generated by the model, and 3) explore the potential of the methodology as what-if analysis tool by addressing a number of problem variants.

The model was implemented as a mixed integer linear program using Java and CPLEX 12.7.0 (single thread). Experiments were performed on a computer Intel(R) Xeon(R) CPU E5-2609 v2 @ 2.50GHz with 8 CPUs and 128 GB de RAM.

An extensive set of instances were generated based on data of a large North American railroad, reflecting realistic problem characteristics and actual practice. The instances include 192 terminals, 519 trains, 5264 demands (OD commodities), 20-, 40- and 53-feet containers, and 40- and 53-feet platforms. 20-feet containers were transformed into 40-feet ones, while 53-feet containers included 43- and 45-feet ones as the same loading rules apply to the three types. The schedule length covers seven days (10080 minutes). We varied the dimensions of the set of potential blocks: a **Complete** (as identified in the following tables) set of 16654 blocks, corresponding to all feasible possibilities (a single restriction only: not more than 24-hour transfer delays); a **Constrained** set of 1929 blocks, where each train has a list of preferred blocks; two sets of intermediate dimensions, **Inter1** and **Inter2**, with 3906 and 7023 blocks, respectively, obtained by restricting the permitted number of transfers or the maximum transfer delay. We performed experiments for the **No-split** demand case, as well as for **Split** demand without any penalty, with **Low penalty** and **High penalty** costs.

Table I displays the results obtained in terms of optimality gaps (in %) after 3 hours and, in between parentheses, 24 hours of CPU time, for the four block sets and the demand-handling scenarios. Not surprisingly, the larger the number of potential blocks, the more difficult it gets to solve the scheduled block service network design problem, pointing to future developments of tailor-made solutions, including dynamic block generation and matheuristics, to address even larger instances. Yet, remarkably good solutions are obtained relatively fast, particularly when practice-oriented problem settings are considered. This is also an indication that, in actual operations, it is not worth letting the software run for very long as proving convergence, which may take a very long time, is not necessarily what is sought after in practice. The next experimental results are displayed for the cases of 3 hours CPU time only.

With respect to the handling of demand, one notices that allowing splitting the demand provides significant efficiency increases in all cases. The more realistic settings are solved to optimality, and dramatic efficiency gains are observed for the largest block set considered. Splitting is beneficial even when low to moderate penalties have to be paid, which is an interesting managerial insight for future negotiations with customers. Very high penalties greatly deteriorate efficiency; it is better in this case, to use the model with the no-split option.

Table I: Efficiency - Optimality Gaps (%)

Block set	No split	Split - No penalty	Split - Low penalty	Split - High penalty
Complete (16.6k)	16 (15)	3 (3)	7 (6)	42 (34)
Constrained (2k)	1 (1)	0 (0)	0 (0)	3 (3)
Inter1 (4k)	2 (1)	0 (0)	0 (0)	3 (3)
Inter2 (7k)	10 (9)	2 (2)	5 (4)	30 (21)

Table II: Block Service Network Structure

Block set	No split	Split - No penalty	Split - Low penalty	Split - High penalty
Complete (16.6k)	905 (1.6)	939 (0)	948 (0.3)	932 (2.9)
Constrained (2k)	721 (12.3)	733 (11.5)	728 (11.5)	731 (12.6)
Inter1 (4k)	727 (12)	744(11)	736 (0)	742 (12.2)
Inter2 (7k)	848 (5.1)	881 (4)	862 (4)	877 (6.3)

Table III: Costs Performance

Block set	No split	Split - No penalty	Split - Low penalty	Split - High penalty
Complete (16.6k)	4.5	3.9	4.0	5.7
Constrained (2k)	8.9	8.7	8.7	9.0
Inter1 (4k)	8.7	8.5	8.5	8.8
Inter2 (7k)	5.8	5.3	5.5	7.5

Table II displays the results in terms of block service network structure and performance with respect to demand fulfilment. The figures indicate for each setting (the same as previously), the number of selected blocks and, in parentheses, the percentage of undelivered demand. One observes that, as expected, starting with a large set of potential blocks yields better results in terms of system performance compared to more restrained sets. A larger, and thus more diversified, initial set means a larger selection of blocks with similar cost values, providing the opportunity to better use the available train capacity and, thus, to more largely spread out the demand among these alternative paths and deliver more. Indeed, the percentage of un-delivered demand is significantly lower for the largest set of potential blocks than for the others. As a managerial insight, this result indicates that even when sets of preferred blocks are attached to trains, these sets should be larger, even though the computational effort will also be larger. The results also show the interest of the proposed model as an analysis tool to evaluate train-selection strategies. In all cases, though, the selected block service network is rather stable, the variations in the numbers of selected blocks is significantly smaller than that of the numbers of potential blocks.

The results with respect to the handling of demand are similar to those of the first set of results. Splitting increases the performance in terms of servicing the demand, as larger blocks service networks and lower percentages of un-delivered demand are observed. High penalties are to be avoided, low-medium-valued ones providing control, i.e., good results, particularly when the number of potential blocks is large, with little performance degradation.

Finally, Table III displays results relative to the cost of the block service network provided by the model. Costs are in millions of Canadian dollars, and were obtained after 3 hours of CPU time. The respective figures after 24 hours of CPU time, not displayed, are a little lower, stating again that proving convergence is a very slow process. One observes significant cost reductions when the model may choose among a large number of potential blocks and even larger when demand splitting is permitted. The cost-related results confirm all the previous conclusions relative to the split/no-split and low/high penalty costs policies.

3.5 Conclusions

We addressed the tactical block planning problem of intermodal rail and proposed a new continuous-time, three-layer service network design model that explicitly addresses the challenging characteristics of intermodal transportation. We performed an extensive experimentation studying the case of a major North American railroad. The results showed the interest of the methodology as one can address realistic instances with commercial software, obtaining good solutions within acceptable computing times. The analysis also provided several managerial insights, e.g., the possibility to split the demand significantly increases the efficiency of the procedure and the quality of the solution with respect to total cost and undelivered demand; high penalties for splitting demand greatly deteriorate efficiency and solution quality, while low-medium ones provides control with little performance degradation; the larger the number of possible blocks, the more blocks are selected and, then, the better trains are used and demand is serviced for a better economic performance of the system.

Chapter 4

Conclusion

Demand for intermodal transporation has seen an important growth over the past decades and has become a fundamental component of global supply chains. Railroads constitute a crucial element of intermodal transportation, with millions of commodities transported every year. Efficient railroad activities are a key to satisfy customers' demand, and require an appropriate planning of operations and resources.

In this thesis, we addressed the tactical block planning problem arising at intermodal rails. Prior to this work, there were only a few publications addressing the traditional train blocking problem, where demand is expressed in terms of cars to be moved and blocked. To best of our knowledge, no work has been done for intermodal transportation, where demand is expressed in terms of containers to be moved from a given origin terminal to a given destination terminal. The thesis contributes to fill this gap in literature, proposing a blocking model that considers several types of containers and railcars, integrating the container-to-car assignment. The model deals with the aforementioned assignment, which is particularly challenging because of the multitude of container and car types existing in North America, as well as the multiplicity of loading units, obliging the explicit integration of the loading part at tactical level.

We proposed a continuous-time, three-layer service network design model to explicitly deal with the challenging characteristics of intermodal transportation. The research is based on a North American use case. As illustrated in Chapter 3, we developed a Mixed Integer Linear Programming (MILP) formulation to address the problem, with the objective of minimizing the total transportation cost composed by block selection, operation costs, and handling demand cost.

We performed an extensive experimental study. The results showed that one can address realistic instances with a commercial MIP solver, obtaining good-quality solutions within acceptable computing time. The analysis also provided several managerial insights, e.g., the possibility to split the demand significantly increases the efficiency of the procedure and the quality of the solution with respect to total cost and undelivered demand; high penalties for splitting demand greatly deteriorate efficiency and solution quality, while low-medium ones provides control with little performance degradation; the larger the number of possible blocks, the more blocks are selected and, then, the better trains are used and demand is serviced for a better economic performance of the system.

Future research can be dedicated to blocking and rail car car fleet management as well as studying the case when containers can be reloaded en-route.

Bibliography

Ahuja, R., C. B. & Şahin, G. (2014), 'Network models in railroad planning and scheduling', `http://dx.doi.org/10.1287/educ.1053.0013`. [Online: Accessed on 14 October 2014].

Ahuja, R. K., Jha, K. C. & Liu, J. (2007), 'Solving real-life railroad blocking problems', *Interfaces* **37**(5), 404–419.

Barnhart, C., Jin, H. & Vance, P. H. (2000), 'Railroad blocking: A network design application', *Operations Research* **48**(4), 603–614.

Bodin, L. D., Golden, B. L., Schuster, A. D. & Romig, W. (1980), 'A model for the blocking of trains', *Transportation Research Part B: Methodological* **14**(1), 115–120.

Boysen, N., Fliedner, M., Jaehn, F. & Pesch, E. (2013), 'A survey on container processing in railway yards', *Transportation Science* **47**(3), 312–329.

Carlo, H. J., Vis, I. F. & Roodbergen, K. J. (2014), 'Transport operations in container terminals: Literature overview, trends, research directions and classification scheme', *European Journal of Operational Research* **236**(1), 1–13.

Crainic, T., Ferland, J.-A. & Rousseau, J.-M. (1984), 'A tactical planning model for rail freight transportation', *Transportation science* **18**(2), 165–184.

Crainic, T. G. & Kim, K. H. (2007), 'Intermodal transportation', *Handbooks in Operations Research and Management Science* **14**(8), 467–537.

Crainic, T. & Rousseau, J.-M. (1986), 'Multicommodity, multimode freight transportation: A general modeling and algorithmic framework for the service network design problem', *Transportation Research Part B: Methodological* **20**, 225–242.

IAPH (2015), ‘World port data 2015’, International Assocaciation of Ports and Harbors, `http://www.iaphworldports.org/statistics`. [Online; accessed 20-September-2016].

Jha, K. C., Ahuja, R. K. & Şahin, G. (2008), ‘New approaches for solving the block-to-train assignment problem’, *Networks* **51**(1), 48–62.

Macharis, C. & Bontekoning, Y. M. (2004), ‘Opportunities for or in intermodal freight transport research: A review’, *European Journal of operational research* **153**(2), 400–416.

Mantovani, S., Morganti, G., Umang, N., Crainic, T. G., Frejinger, E. & Larsen, E. (2017), ‘The load planning problem for double-stack intermodal trains’, *European Journal of Operational Research* **267**(1), 107–119.

Newton, H. N., Barnhart, C. & Vance, P. H. (1998), ‘Constructing railroad blocking plans to minimize handling costs’, *Transportation Science* **32**(4), 330–345.

Steenken, D., Voß, S. & Stahlbock, R. (2004), ‘Container terminal operation and operations research-a classification and literature review’, *OR spectrum* **26**(1), 3–49.

Vis, I. F. & De Koster, R. (2003), ‘Transshipment of containers at a container terminal: An overview’, *European journal of operational research* **147**(1), 1–16.

Zhu, E., Crainic, T. G. & Gendreau, M. (2014), ‘Scheduled service network design for freight rail transportation’, *Operations Research* **62**(2), 383–400.